Introduction to Linked Open Data

Tutorial

DC-2013 Conference

2nd September, 2013, Lisbon, Portugal

Ivan Herman, W3C

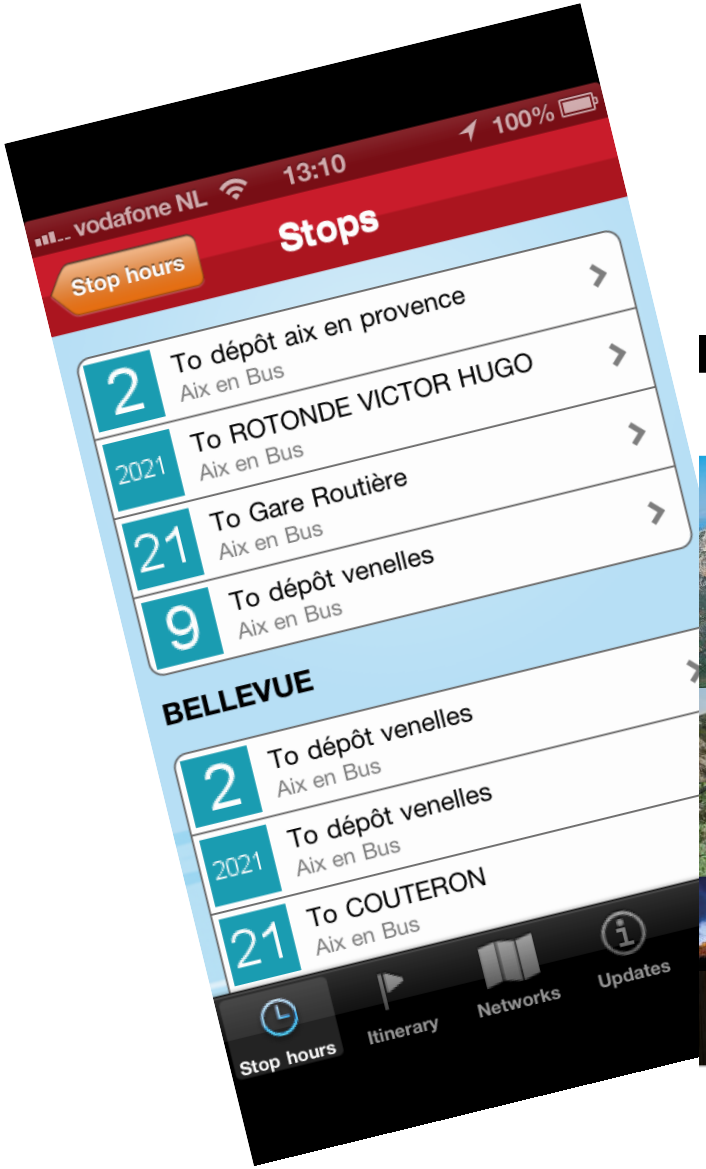
DC-2013 Linking to the Future Lisbon

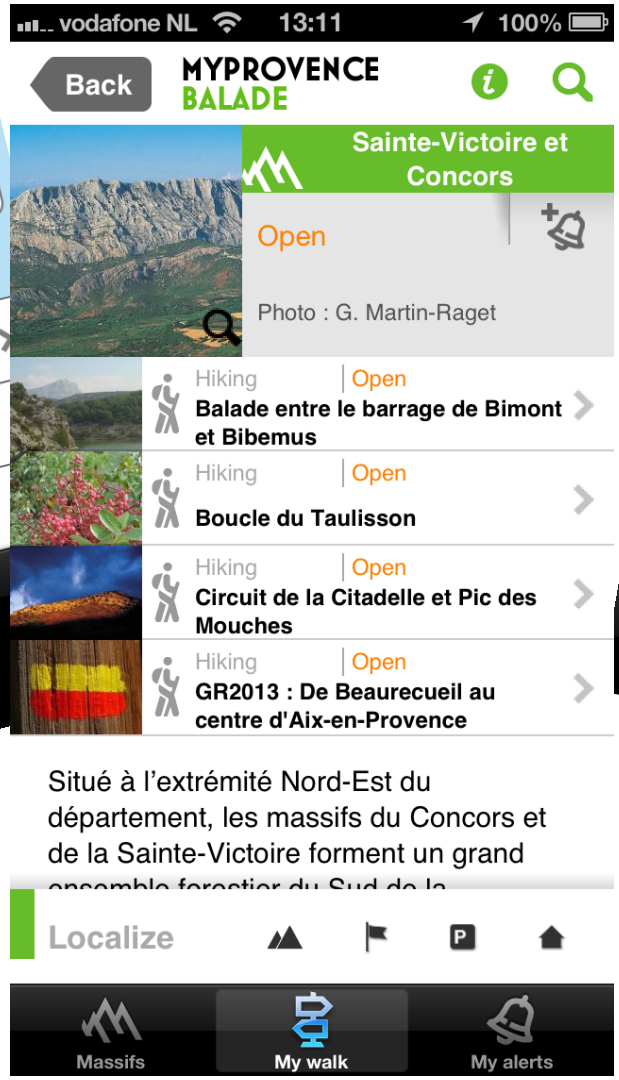
W3C World Wide Web consortium

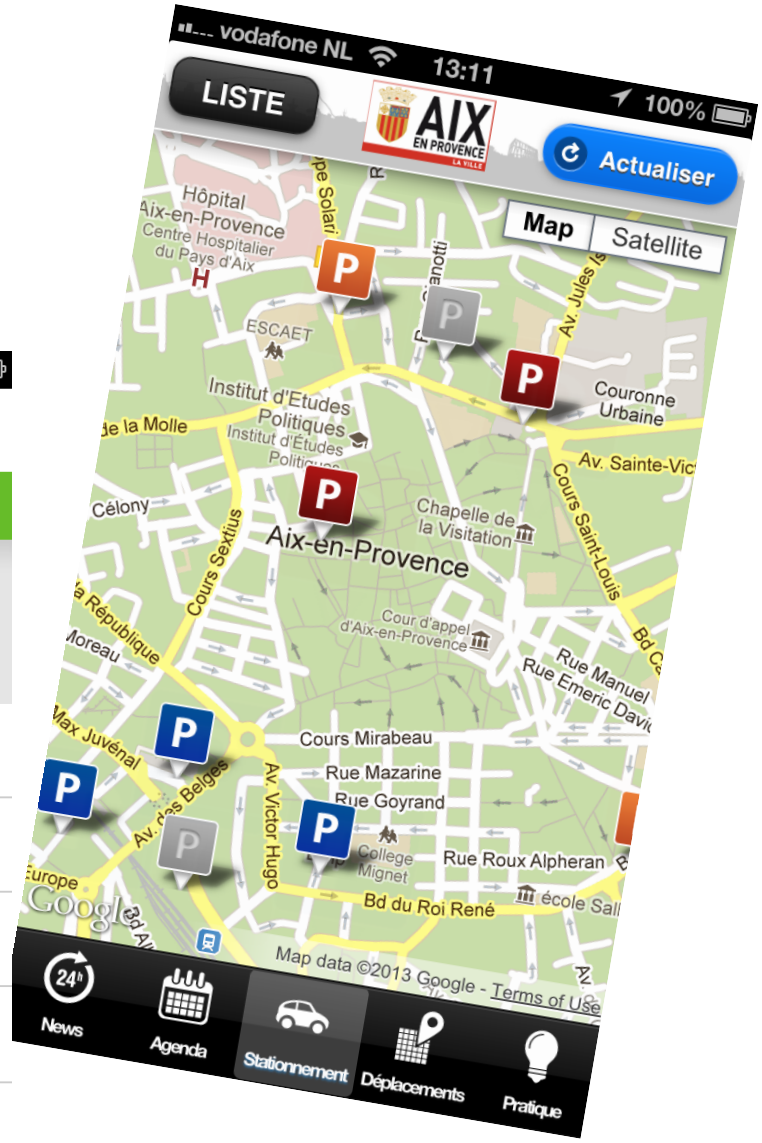
These slides are also available on the Web:

http://www.w3.org/2013/Talks/0902-Lisbon-IH/

Introduction


vodafone NL
13:10
100%
Stop hours
Stops
2
To dépôt aix en provence
Aix en Bus
2021
To ROTONDE VICTOR HUGO
Aix en Bus
21
To Gare Routière
Aix en Bus
9
To dépôt venelles
Aix en Bus
BELLEVUE
2
To dépôt venelles
Aix en Bus
2021
To dépôt venelles
Aix en Bus
21
To COUTERON
Aix en Bus
Stop hours
Itinerary
Networks
Updates


vodafone NL
13:11
100%
Back
MYPROVENCE
BALADE
Sainte-Victoire et Concors
Open
Photo : G. Martin-Raget
Hiking
Open
Balade entre le barrage de Bimont et Bibemus
Hiking
Open
Boucle du Taulisson
Hiking
Open
Circuit de la Citadelle et Pic des Mouches
Hiking
Open
GR2013 : De Beaurecueil au centre d'Aix-en-Provence
Situé à l'extrémité Nord-Est du département, les massifs du Concors et de la Sainte-Victoire forment un grand
Localize
Massifs
My walk
My alerts


vodafone NL
13:11
100%
LISTE
AIX EN PROVENCE
Actualiser
Map
Satellite
Hôpital Aix-en-Provence
Centre Hospitalier du Pays d'Aix
ESCAET
Institut d'Etudes Politiques
Couronne Urbaine
Av. Sainte-Vic
Chapelle de la Visitation
Aix-en-Provence
Cours Sextius
Cours Saint-Louis
Cour d'appel d'Aix-en-Provence
Rue Manuel
Rue Emeric Davi
Cours Mirabeau
Rue Mazarine
Rue Goyrand
Av. Victor Hugo
College Mignet
Rue Roux Alpheran
Bd du Roi René
Map data ©2013 Google - Terms of Use
News
Agenda
Stationnement
Déplacements
Pratique

Home | Data.gov

www.data.gov

My Mercurial | LocalData | TR | TR List | 2013 | Private | Social | Mailing lists | SW | Python | RDFa it! | Bookmarklets | Mob. Highlighter

An Official Web Site of the United States Government | Change Text Size | SHARE

DATA.GOV
EMPOWERING PEOPLE

Login

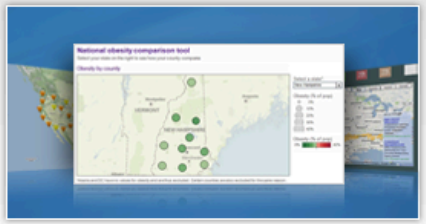
Search Data.gov | SEARCH

HOME | ABOUT | DATA | METRICS | OPEN GOVERNMENT | BLOGS | COMMUNITIES

Latest Datasets

- DOE Green Energy
- DOE Patents
- ScienceCinema
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- Gravesite locations of Veterans and...
- SciTech Connect

DATA AND TOOLS



- 373,029 raw and geospatial datasets
- 1,209 data tools
- 350 apps
- 137 mobile apps
- 172 agencies and subagencies
- Suggest a dataset

COMMUNITIES

Come explore, discuss, meet others in the same field, and develop the data and apps in the community that you care about. Join in the discussions by going to communities, some of which are listed below, that interest you.

EDUCATION | ENERGY
BUSINESSUSA | AGRICULTURE
OCEAN | SAFETY

OPEN GOVERNMENT

View our government at work and find opportunities to participate.

- Open Government Initiative
- Open Government Platform
- Recovery.gov
- USASpending.gov
- We the People

Join the conversation

LEARN

SEMANTIC WEB

DEVELOPERS CORNER

Le Pavillion D'Or (Open Library)

openlibrary.org/books/OL11039953M/Le_Pavillion_D'Or

My Mercurial | LocalData | TR | TR List | 2013 | Private | Social | Mailing lists | SW | Python | RDFa it! | Bookmarklets | Mob. Highlighter

SUBJECTS

AUTHORS | ADD A BOOK

LISTS | RECENTLY | HELP

OPEN LIBRARY

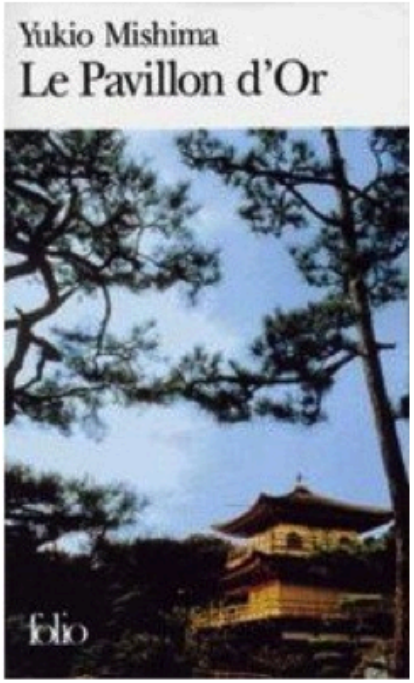
One web page for every book.

Show only eBooks

Cover of: Le Pavillion D'Or by Yukio Mishima

1 edition of Le Pavillion D'Or by Yukio Mishima • Add edition?

Last edited by Ivan Herman
584 milliseconds ago | History



Manage Covers

Le Pavillion D'Or

Published **October 1, 1975** by French & European Pubns.

The Physical Object

Format	Paperback

ID Numbers

Open Library	OL11039953M
ISBN 10	0785940391
ISBN 13	9780785940395
Goodreads	756054

Read

No readable version available.

Borrow

Physical copy, local WorldCat

Buy

Alibris
Amazon
AbeBooks
Biblio.com
Book Depository
Powells

Lists

Connecting to www.archive.org...

Take a specific example

The Music site of the BBC

BBC – Music – Eric Clapton

www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

BBC | Sign in | News | Sport | Weather | iPlayer | TV | Radio | More... | Search

iPlayer Radio | WHAT'S NEW? | Stations | Categories | Programmes | Favourites

MUSIC

HOME | SHOWCASE | GENRES

Search By Artist...

Eric Clapton

Born 30 March 1945.

PLAYED MOST ON BBC RADIO 2

6 MUSIC LIVE HOUR | BBC RADIO 6 MUSIC

Eric Clapton - Archive Concert (1986)

The Eric Clapton band recorded live in concert in Dallas, Texas in 1986.

Share This Page

10 so far

Share

More BBC Music Highlights

Latest Tracks Played On The BBC

Wonderful Tonight

BBC RADIO 2 | STEVE WRIGHT'S SUNDAY LOVE SONGS

The Music site of the BBC

BBC – Music – Eric Clapton

www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

helped reggae reach a mass market. Two of his most popular recordings were "Layla", recorded by Derek and the Dominos, another band he formed, and Robert Johnson's "Crossroads", recorded by Cream. Following the death of his son Conor in 1991, Clapton's grief was expressed in the song "Tears in Heaven", which featured in his Unplugged album.

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information

LINKS

Official homepage at ericclapton.com

Fanpage at whereseric.com

Youtube at youtube.com/user/ericclapton

Twitter at twitter.com/EricClaptonNews

MySpace at myspace.com/ericclapton

Wikipedia article on Eric Clapton

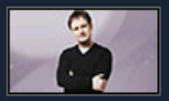
Last.fm page on Eric Clapton

Discogs at discogs.com/artist/Eric Clapton

MusicBrainz entry on Eric Clapton

MEMBER OF Cream (1966-1968), The Yardbirds (1963-1965), Derek and the Dominos, John Mayall & The Bluesbreakers (1965-1966), Blind Faith (1968-1969), The Louisiana Gator Boys

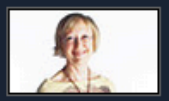
COLLABORATED ON The Dirty Mac (1968)



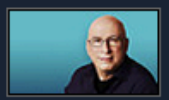
Richard Allinson
BBC Radio 2

Gerry Anderson
BBC Radio Ulster

The Chris Evans Breakfast Show
BBC Radio 2



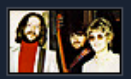
The Late Show with Cherrie McIlwaine
BBC Radio Ulster



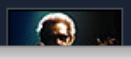
Ken Bruce
BBC Radio 2

Information displayed about artists played on BBC programmes is incomplete at present. Find out more about this artist play count information.

Similar Artists



Derek and the Dominos



J.J. Cale

How to build such a site 1.

- Site editors roam the Web for new facts
  - may discover further links while roaming
- They update the site manually
- And the site gets soon out-of-date ☹

How to build such a site 2.

- Editors roam the Web for new data published on Web sites
- “Scrape” the sites with a program to extract the information
  - Ie, write some code to incorporate the new data
- Easily get out of date again… ☹

How to build such a site 3.

- Editors roam the Web for new data via API-s
- Understand those…
  - input, output arguments, datatypes used, etc
  - hope that the necessary portion of the data is released through the API
- Write some code to incorporate the new data
- Easily get out of date again… ☹

The choice of the BBC

- Use external, public datasets
  - Wikipedia, MusicBrainz, …
- They are available *as data*
  - not API-s or hidden on a Web site
  - data can be extracted using, e.g., HTTP requests or standard queries

In short…

- Use the Web of Data as a Content Management System
- Use the community at large as content editors

And this is no secret…

BBC – Music – Eric Clapton

www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

Last.fm page on Eric Clapton

Discogs at discogs.com/artist/Eric Clapton

MusicBrainz entry on Eric Clapton

MEMBER OF Cream (1966-1968), The Yardbirds (1963-1965), Derek and the Dominos, John Mayall & The Bluesbreakers (1965-1966), Blind Faith (1968-1969), The Louisiana Gator Boys

COLLABORATED ON The Dirty Mac (1968)

PERSONAL RELATIONSHIPS

Married to Pattie Boyd (1979-1989)

Links & information come from MusicBrainz. You can add or edit information about Eric Clapton at musicbrainz.org. Find out more about our use of this data. The BBC is not responsible for the content of external sites.

BBC Reviews

Me And Mr Johnson 2004

REVIEWED BY ROB WEBB

“” So, 35 years on, Clapton is no longer God: he now plays the Devil

Martin Scorsese Presents... 2003

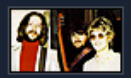
REVIEWED BY MICK FITZSIMMONS

“” ...this hits the spot nicely and underlines just why Clapton is so revered amongst...

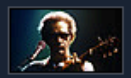
More Eric Clapton releases at MusicBrainz »

information.

Similar Artists



Derek and the Dominos



J.J. Cale

John Mayall

Stevie Ray Vaughan

We use data provided by The Echo Nest to recommend similar artists. Find out more about this information.

And this is no secret…

BBC – Music – Eric Clapton

www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

helped reggae reach a mass market. Two of his most popular recordings were "Layla", recorded by Derek and the Dominos, another band he formed, and Robert Johnson's "Crossroads", recorded by Cream. Following the death of his son Conor in 1991, Clapton's grief was expressed in the song "Tears in Heaven", which featured in his Unplugged album.

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information

LINKS

Official homepage at ericclapton.com

Fanpage at whereseric.com

Youtube at youtube.com/user/ericclapton

Twitter at twitter.com/EricClaptonNews

MySpace at myspace.com/ericclapton

Wikipedia article on Eric Clapton

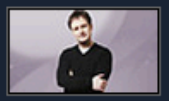
Last.fm page on Eric Clapton

Discogs at discogs.com/artist/Eric Clapton

MusicBrainz entry on Eric Clapton

MEMBER OF Cream (1966-1968), The Yardbirds (1963-1965), Derek and the Dominos, John Mayall & The Bluesbreakers (1965-1966), Blind Faith (1968-1969), The Louisiana Gator Boys

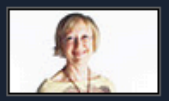
COLLABORATED ON The Dirty Mac (1968)



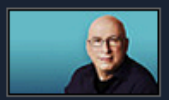
Richard Allinson
BBC Radio 2

Gerry Anderson
BBC Radio Ulster

The Chris Evans Breakfast Show
BBC Radio 2



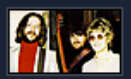
The Late Show with Cherrie McIlwaine
BBC Radio Ulster



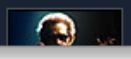
Ken Bruce
BBC Radio 2

Information displayed about artists played on BBC programmes is incomplete at present. Find out more about this artist play count information.

Similar Artists



Derek and the Dominos



J.J. Cale

Data on the Web

- There are more an more data on the Web
  - government data, health related data, general knowledge, company information, flight information, restaurants,…
- More and more applications rely on the availability of that data

But… data are often in isolation, “silos”

 Photo credit “nepatterson”, Flickr

Imagine…

- A “Web” where
  - documents are available for download on the Internet
  - but there would be no hyperlinks among them

And the problem *is* real…

CoCoDat - Collation of Cortical Data - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.cocom

CoCoMac DATABASES ORT EXAMPLES

CoCoDat: Collation of Cortical [s
microcircuitry] Data

CoCoDat is a microcircuitry database that
published experimental reports. The data
and cellular compartment), as well as the

- Morphology
- Firing properties
- Ionic currents
- Ionic conductances
- Synaptic currents
- Connectivity

The database is available for download u
data tables but also a Search Board with p
manual or automatic relaxation of the sea

- Brain region
- Layer
- Neuron type

http://www.cocomac.org/cocodat/catalyzer/ind

Cell Centered Database - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ccdb.ucsd.edu/sand/main?event=gallery&action=show&dpl=y

red database

Cell Centered Database™
National Center for Microscopy and Imaging Research
Gallery

Data | Search | Gallery | Dictionary | Publications | MyCCDB | Data Download | Contact us | Help

2D image | Reconstruction | Segmentation | Animation

y 23
D:
e 071202tjd
e:
D: P1188
ct Brain Maps
e:
ct Tom Deerinck
r:
nt BioRad RTS 2000MP
d: Multiphoton
n: montage of cerebellar cortex triple labeled for Hoescht stain (blue), IP3 receptor (green) and GFAP (red)

logged out

NeuronDB - Thalamic relay neuron - Overview (A) () - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://senselab.med.yale.edu

senselab

NeuronDB

Back

Thalamic relay neuron

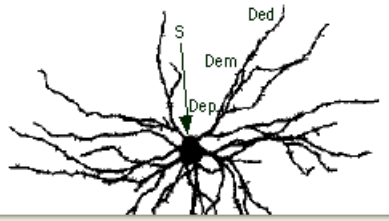
Mode: Overview | Data/Search | plus Connectivity | plus Classical References/Notes | Models

Region: Distal equivalent dendrite | Middle equivalent dendrite | Proximal equivalent dendrite | Soma | Axon hillock | Axon fiber | Axon terminal | All Compartments

Properties: Receptors | Channels | Transmitters | All Properties

Interoperation: Gene and Chromosome | Experimental Data (neurodatabase.org) | Microscopy Data (CCDB)

Neuron type: principal
Organism: Vertebrates



1. Equivalent dendrite — Show other
2. Distal equivalent dendrite — Show other
3. Middle equivalent dendrite — Show other
4. Proximal equivalent dendrite — Show other
5. Soma — Show other

Done

logged out

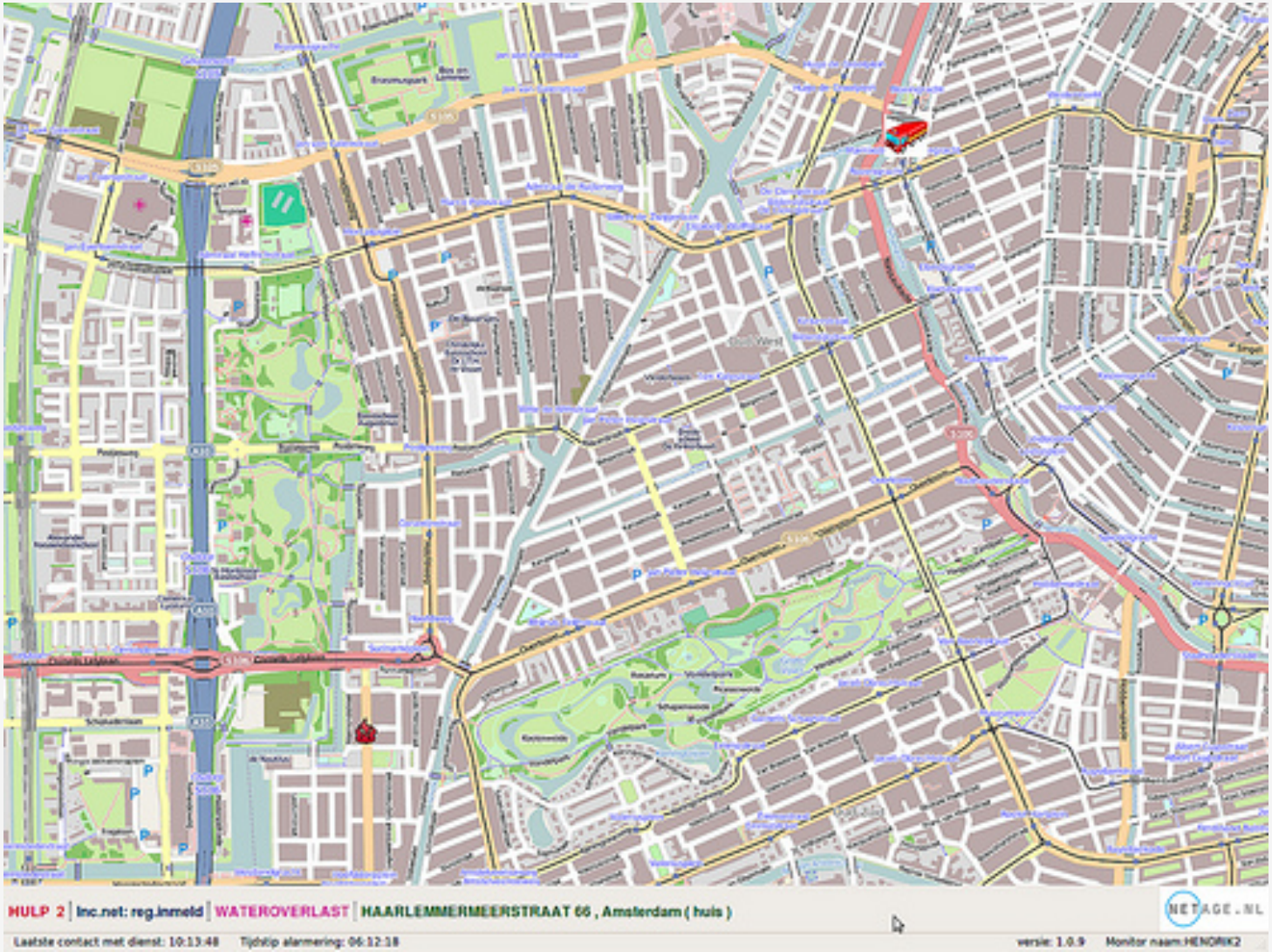
Data on the Web is not enough…

- We need a proper infrastructure for a real *Web of Data*
  - data is available on the Web
    - accessible via standard Web technologies
  - data are interlinked over the Web
  - i.e., data can be *integrated* over the Web
- This is where Linked Data come in

I.e.,… connect the silos

 Photo credit "kxlly", Flickr

Example: Amsterdam fire brigade routing



- Find the best possible route from the station to the fire
  - e.g., where are the roadblocks?
- Use and integrate available city data
- Also: republish the structured data for others to use!

Courtesy of Bart van Leeuwen, Amsterdam Fire Service, The Netherlands

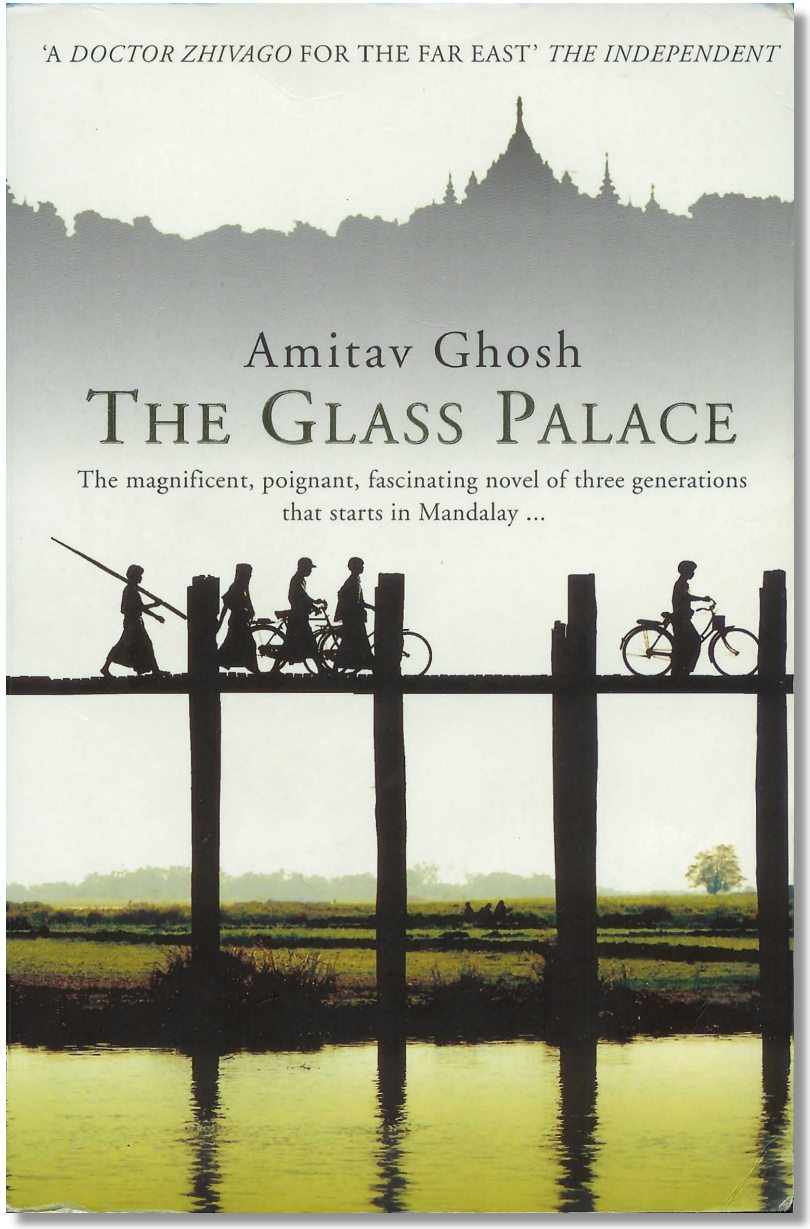
In what follows…

- We will use a simplistic example to introduce the main technical concepts

The rough structure of data integration

- Map the various data onto an abstract data representation
  - make the data independent of its internal representation…
- Merge the resulting representations
- Start making, e.g., queries on the whole!
  - queries not possible on the individual data sets

We start with a book...



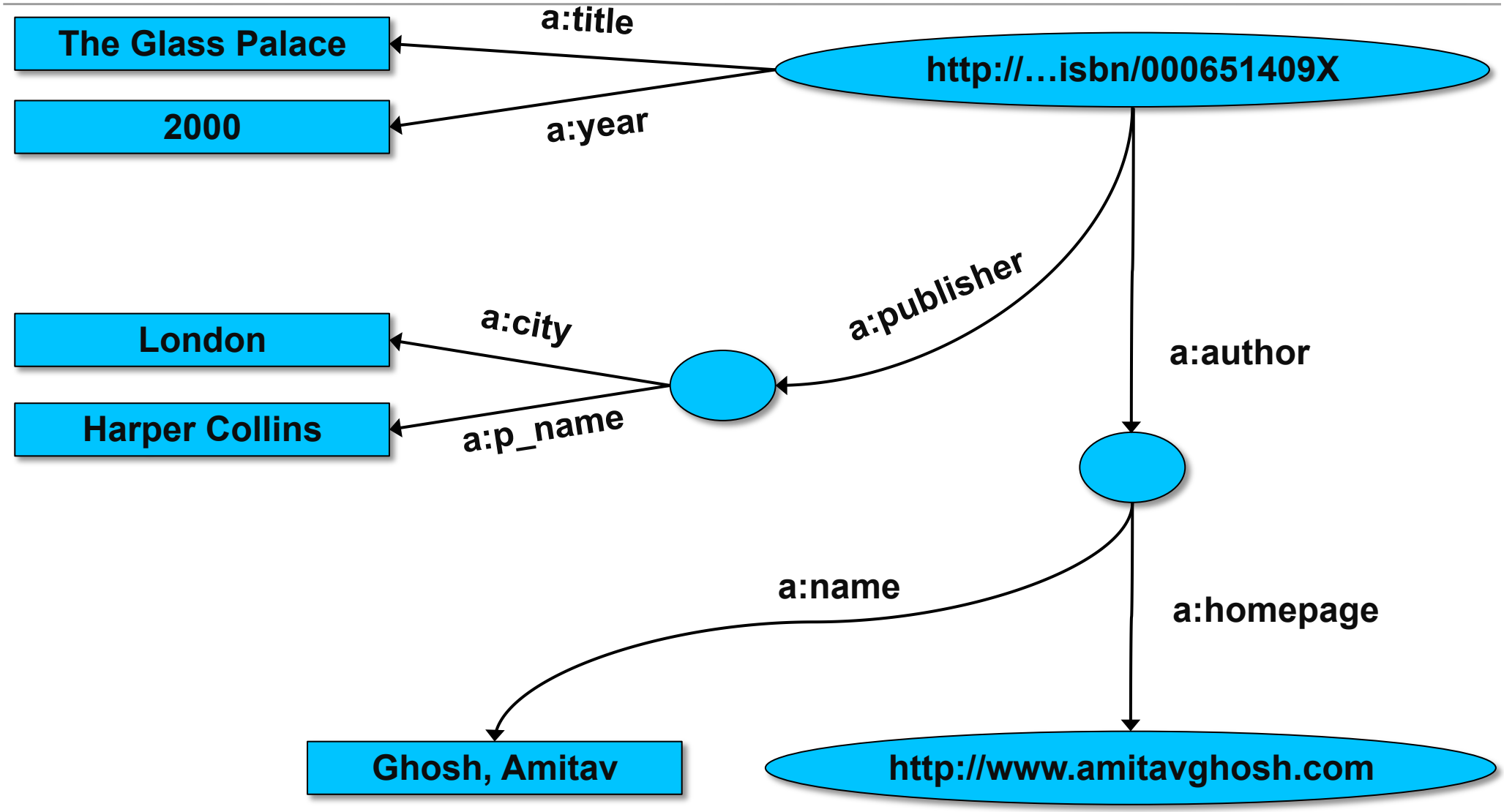
A simplified bookstore data (dataset “A”)

ISBN	Author	Title	Publisher	Year
0006511409X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID	Publisher’s name	City
id_qpr	Harper Collins	London

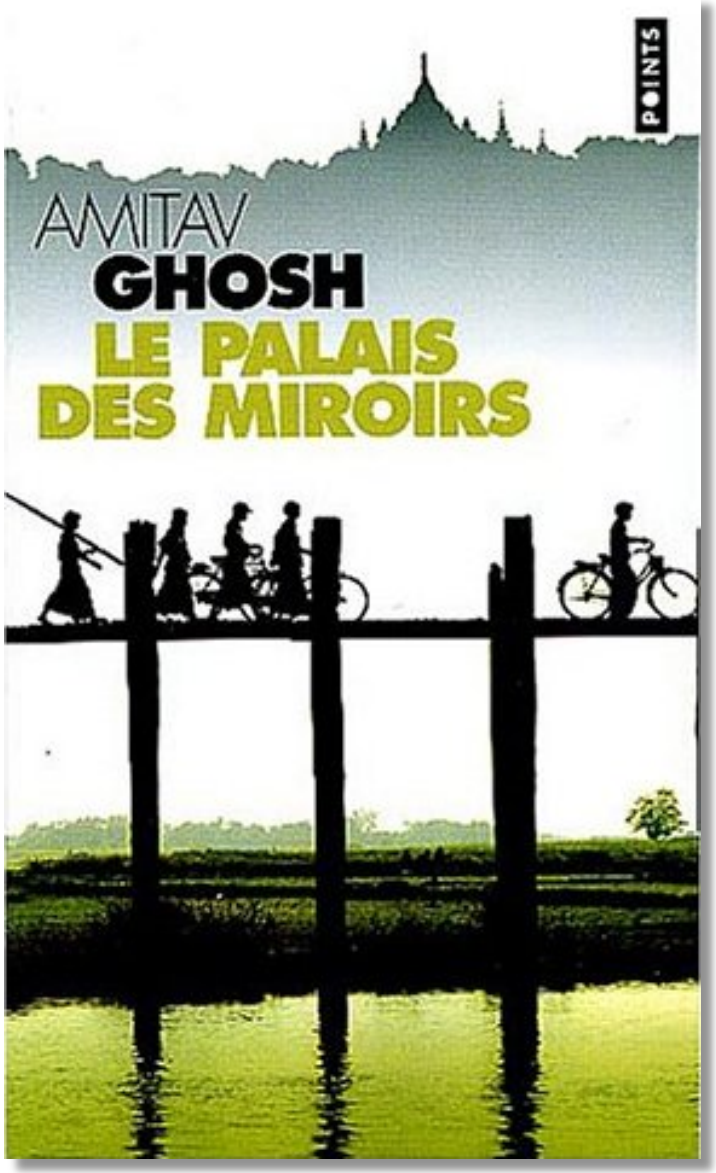
1st: export your data as a set of *relations*



Some notes on the exporting the data

- Relations form a graph
  - the nodes refer to the “real” data or contain some literal
  - how the graph is represented in machine is immaterial for now

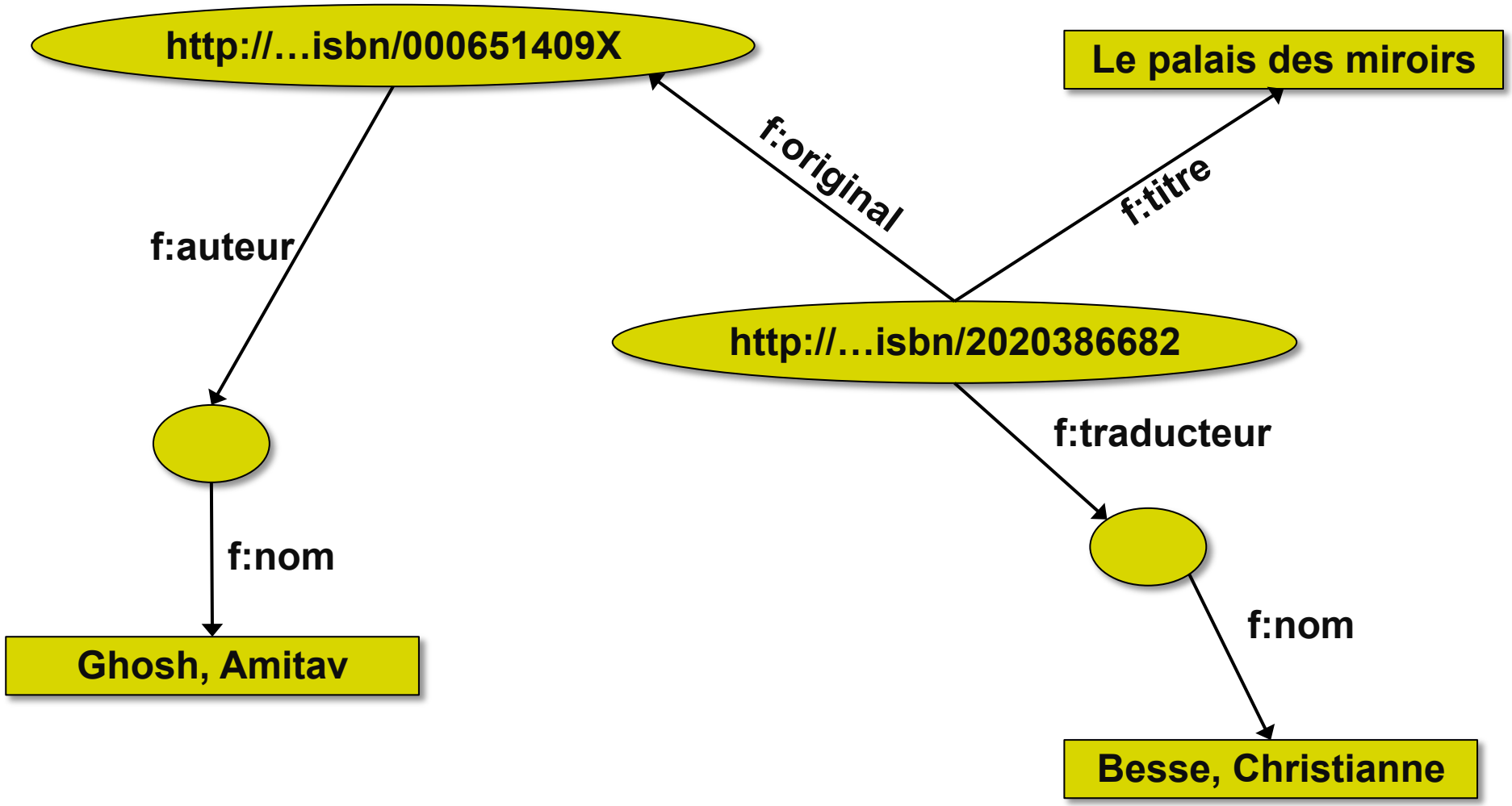
Same book in French…



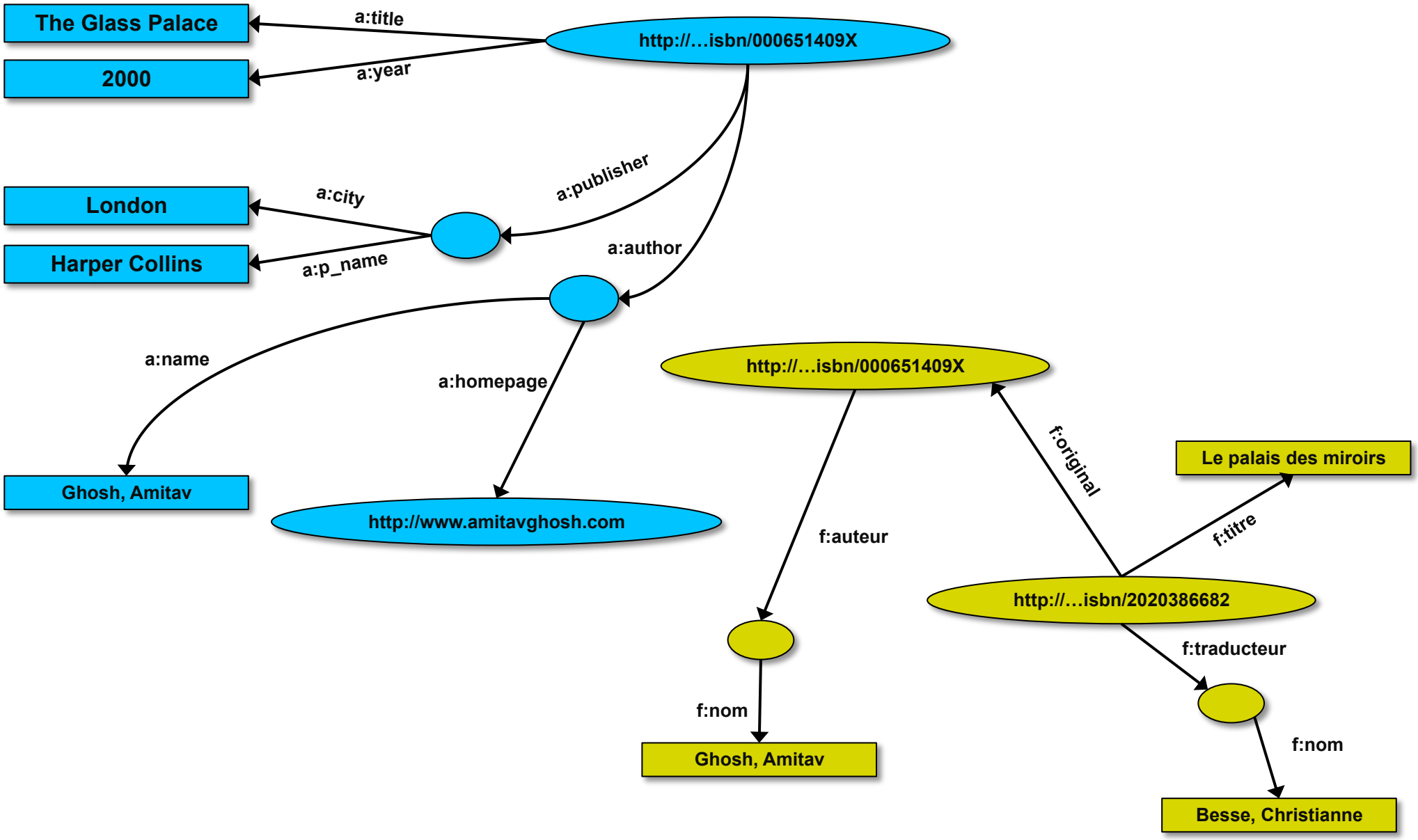
Another bookstore data (dataset “F”)

	A	B	C	D
1	ID	Titre	Traducteur	Original
2	ISBN 2020286682	Le Palais des Miroirs	$A12$	ISBN 0-00-6511409-X
3				
4				
5				
6	ID	Auteur		
7	ISBN 0-00-6511409-X	$A11$		
8				
9				
10	Nom			
11	Ghosh, Amitav			
12	Besse, Christianne			

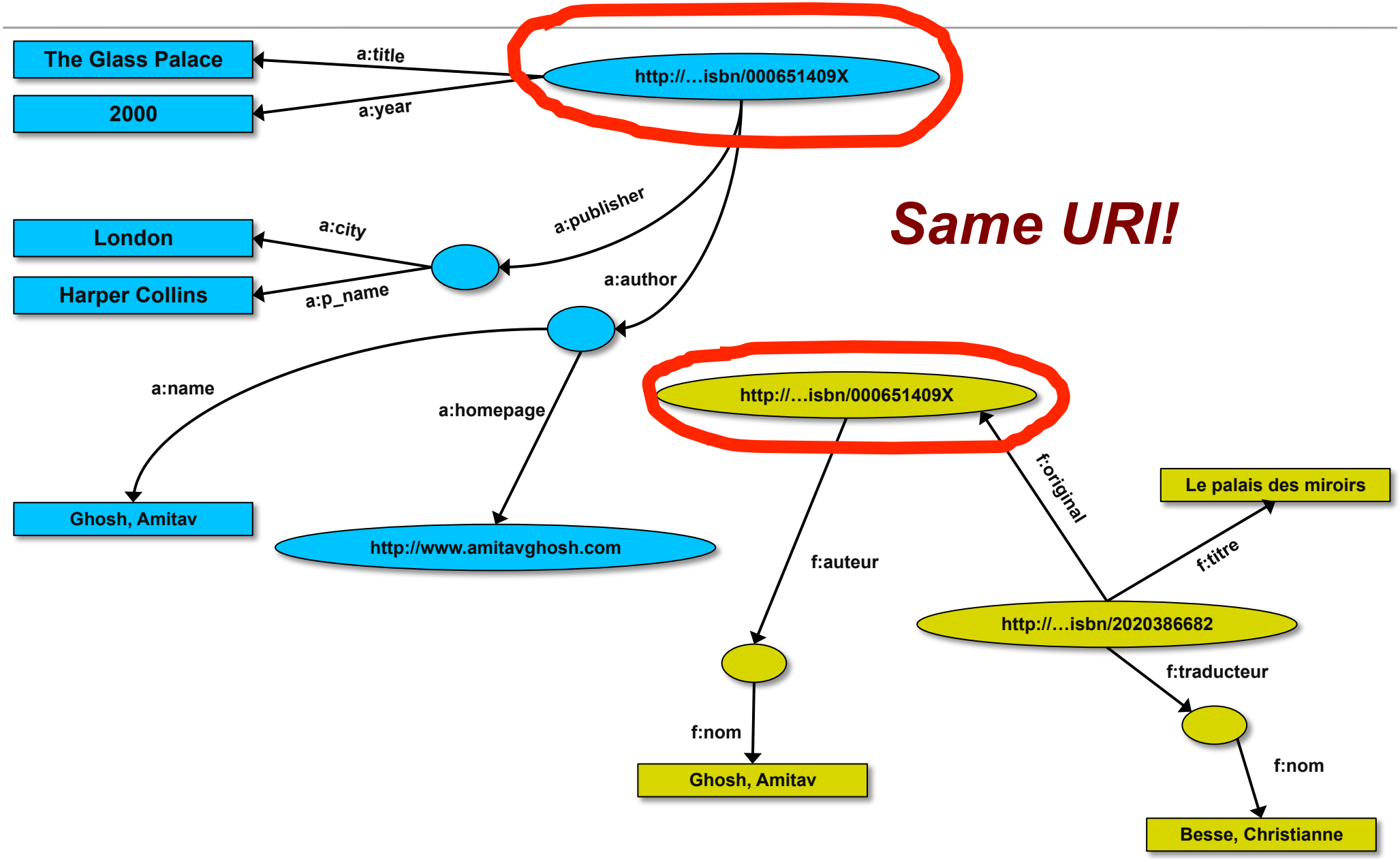
2nd: export your second set of data



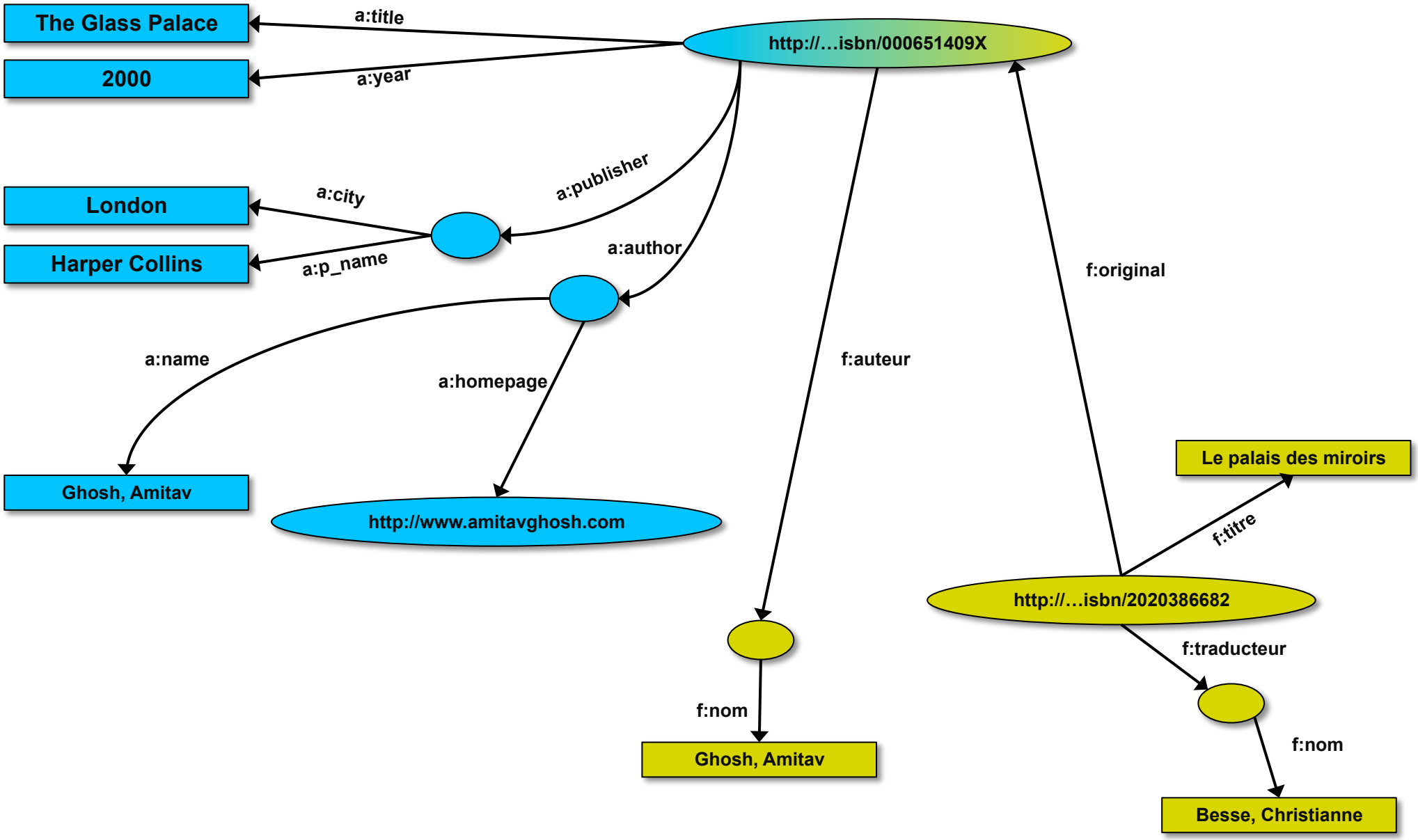
3rd: start merging your data



3rd: start merging your data (cont)

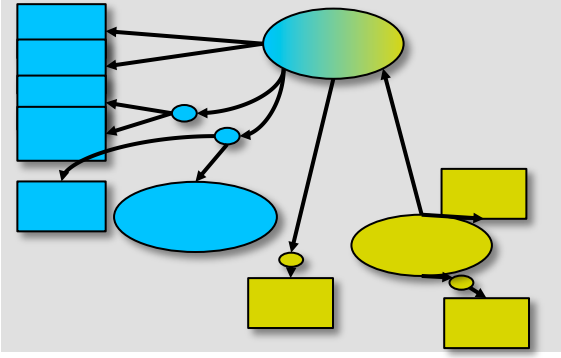


3rd: start merging your data



Start making queries…

- User of data “F” can now ask queries like:
  - “give me the title of the original”
    - well, … « donnes-moi le titre de l’original »
- This information is not in the dataset “F”…
- …but can be retrieved by merging with dataset “A”!

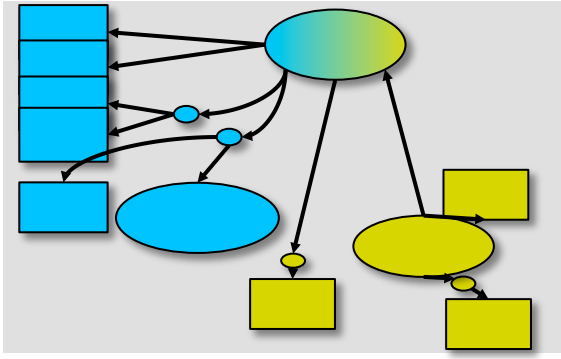


However, more can be achieved…

- We “feel” that a:author and f:auteur should be the same
- But an automatic merge doest not know that!
- Let us add some extra information to the merged data:
  - a:author same as f:auteur
  - both identify a “Person”
  - a term that a community may have already defined:
    - a “Person” is uniquely identified by his/her name and, say, homepage
    - it can be used as a “category” for certain type of resources

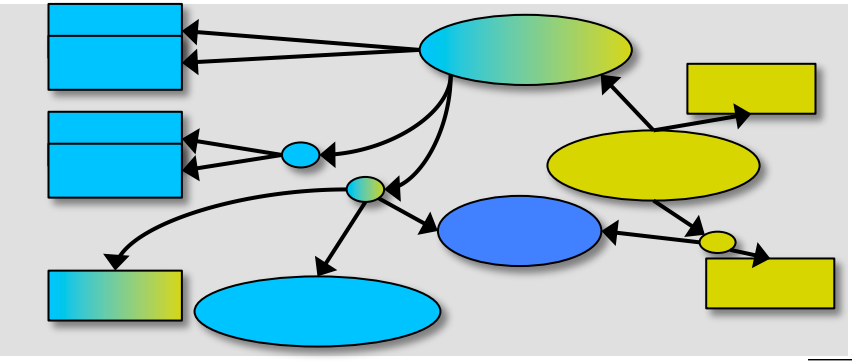
3rd revisited: use the extra knowledge

The Glass Palace
a:title
2000
a:year
http://…isbn/000651409X
f:original
Le palais des miroirs
f:titre
London
a:city
a:publisher
Harper Collins
a:p_name
a:author
f:auteur
http://…isbn/2020386682
r:type
f:traducteur
a:name
f:nom
a:homepage
http://…foaf/Person
r:type
f:nom
Besse, Christianne
Ghosh, Amitav
http://www.amitavghosh.com



Start making richer queries!

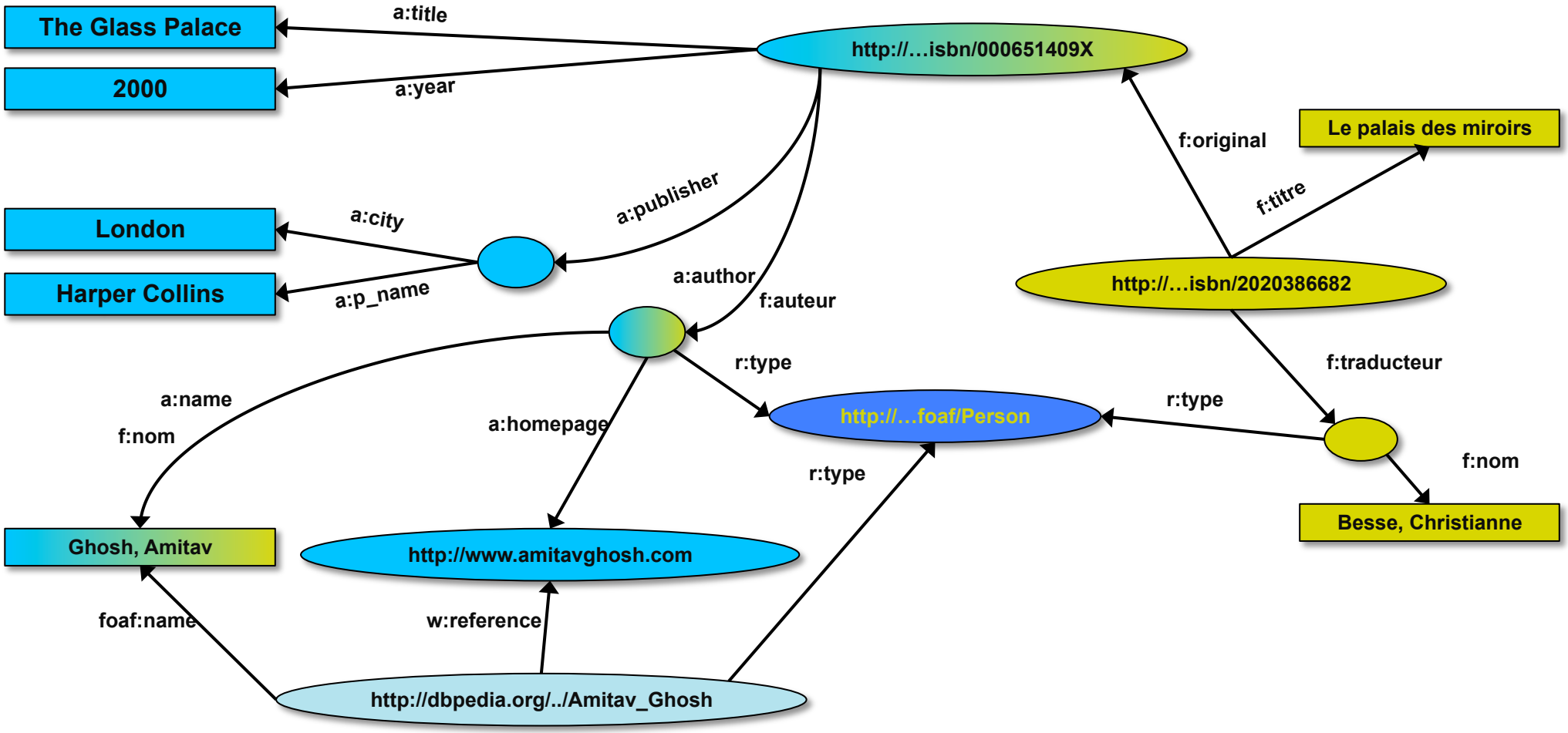
- User of dataset “F” can now query:
  - “donnes-moi la page d’accueil de l’auteur de l’original”
    - well… “give me the home page of the original’s ‘auteur’”
- The information is not in datasets “F” or “A”…
- …but was made available by:
  - merging datasets “A” and datasets “F”
  - adding three simple extra statements as an extra “glue”



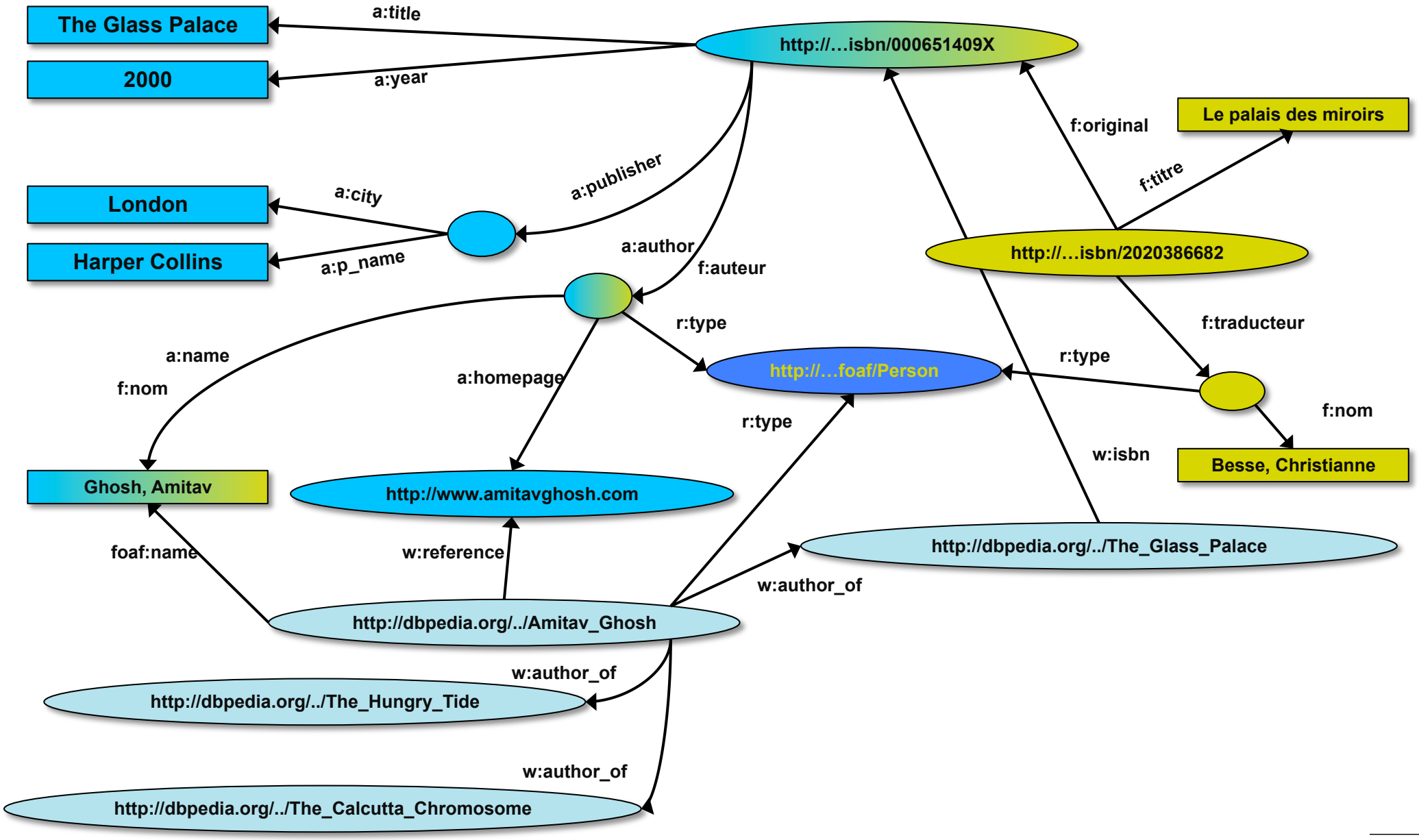
Combine with different datasets

- Using, e.g., the “Person”, the dataset can be combined with other sources
- For example, data in Wikipedia can be extracted using dedicated tools
  - e.g., the “dbpedia” project can extract the “infobox” information from Wikipedia already…

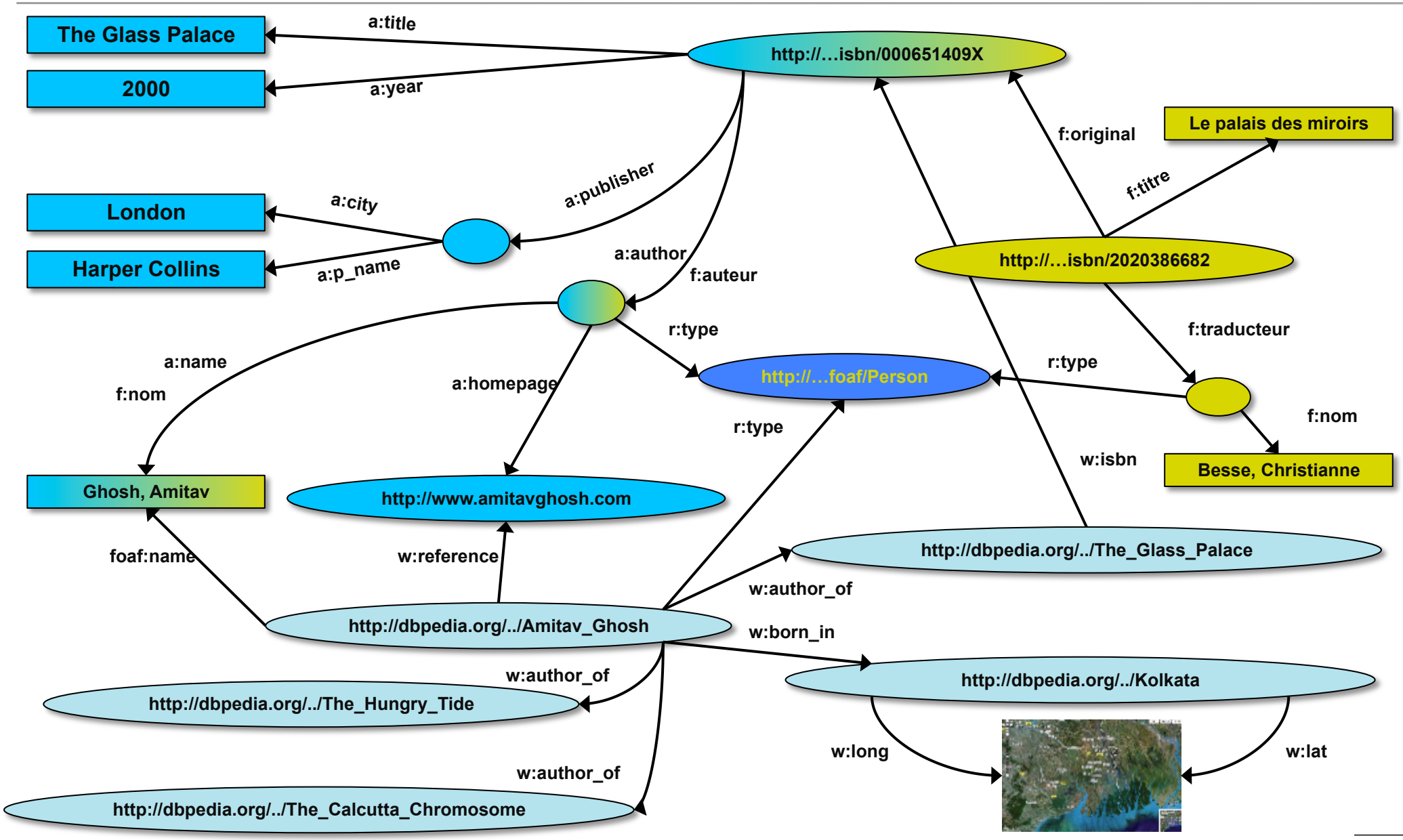
Merge with Wikipedia data



Merge with Wikipedia data



Merge with Wikipedia data



Is that surprising?

- It may look like it but, in fact, it should not be…
- What happened via automatic means is done every day by Web users!
- The difference: a bit of extra rigour so that machines could do this, too

It could become even more powerful

- We could add extra knowledge to the merged datasets
  - e.g., a full classification of various types of library data
  - geographical information
  - etc.
- This is where vocabularies, extra rules, etc., come in
  - vocabularies can be relatively simple and small, or huge, or anything in between…
- Even more powerful queries can be asked as a result

What did we do?

- We identified, uniquely, the pieces of data
- We used URI-s to do that
  - we have the usual Web technologies at our disposal
- De-referencing those URIs usually returns useful information
  - remember the way we could explore the wikipedia/dbpedia data
- We abstracted out of the specific data representation; concentrated on the *links* among data points
- We could also deduce, in some cases, new information

So where is the “Linked Data”? So where is the “Semantic Web”?

- The Semantic Web provides technologies to make such integration possible
- Linked Data is a set of general principles whose realizations are (usually) based on Semantic Web technologies
- Hopefully you get a full picture at the end of the tutorial…

The Basis: RDF

RDF triples

- Let us begin to formalize what we did!
  - we “connected” the data…
  - but a simple connection is not enough… data should be named somehow
  - hence the RDF Triples: a labelled connection between two resources

RDF triples (cont.)

- An RDF Triple (s,p,o) is such that:
  - “s”, “p” are URI-s, ie, resources on the Web; “o” is a URI or a literal
    - “s”, “p”, and “o” stand for “subject”, “property”, and “object”
  - here is the complete triple:

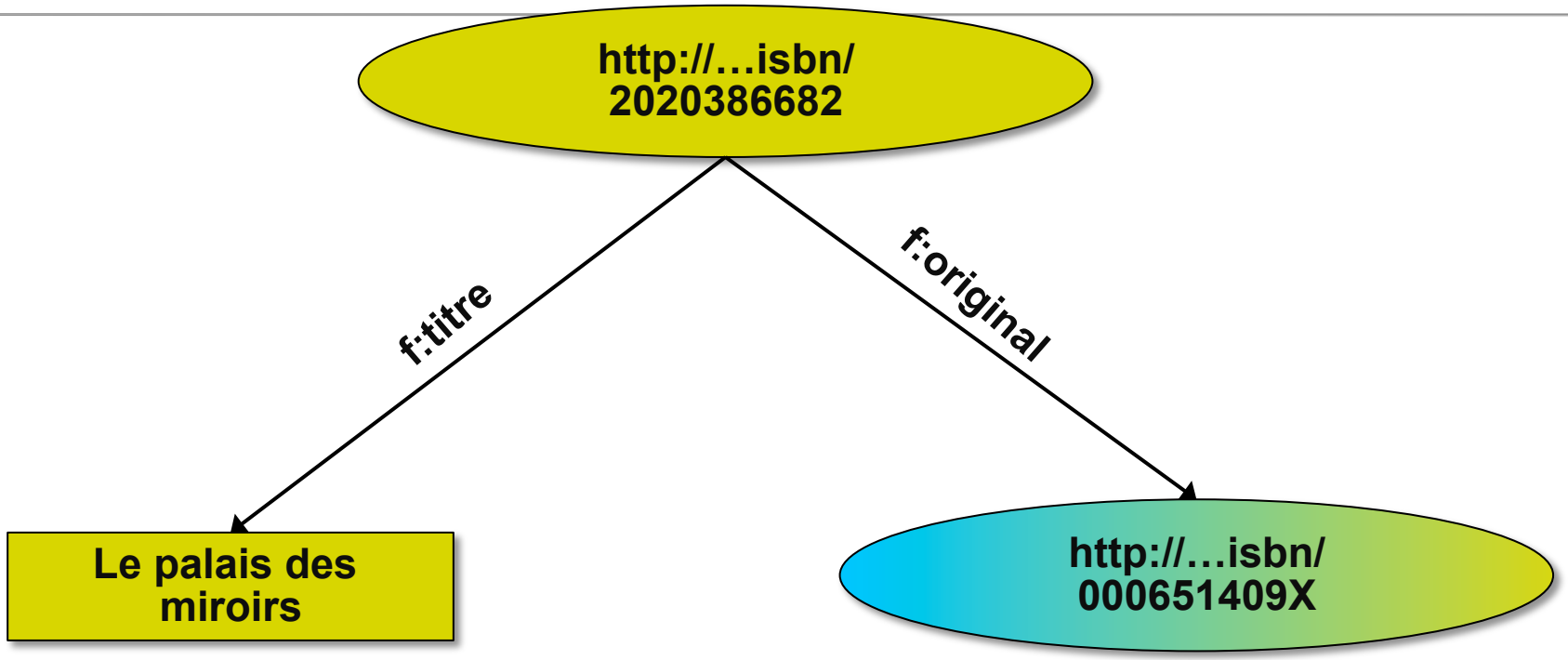
```
(<http://…isbn…6682>, <http://…/original>, <http://…isbn…409X>)
```

- *RDF* is a general model for such triples (with machine readable formats like RDF/XML, Turtle, N3, RDFa, Json, …)

RDF triples (cont.)

- Resources can use *any* URI
  - http://www.example.org/file.html#home
  - http://www.example.org/form?a=b&c=d
  - doi:10.3998/3336451.0004.203
- RDF triples form a directed, labeled graph (the best way to think about them!)
  - the abstract concepts are serialized in XML, Turtle, JSON-LD, …

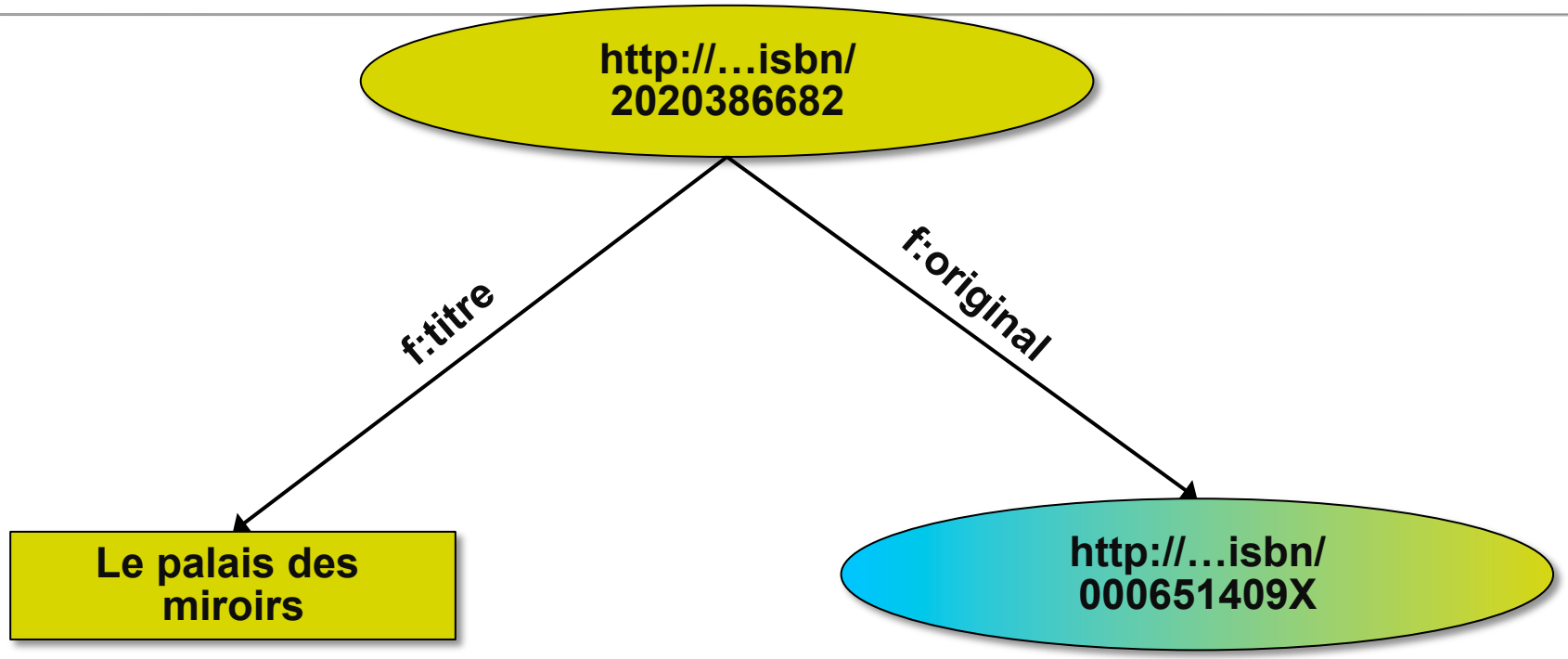
A simple RDF example (in RDF/XML)




```
<rdf:Description rdf:about="http://…/isbn/2020386682">
    <f:titre xml:lang="fr">Le palais des mirroirs</f:titre>
    <f:original rdf:resource="http://…/isbn/000651409X"/>
</rdf:Description>
```

(Note: namespaces are used to simplify the URI-s)

A simple RDF example (in Turtle)




```
<http://…/isbn/2020386682>
    f:titre "Le palais des mirroirs"@fr ;
    f:original <http://…/isbn/000651409X> .
```

RDF in programming practice

- For example, using Python+RDFLib:
  - a “Graph” object is created
  - the RDF file is parsed and results stored in the Graph
  - the Graph offers methods to retrieve:
    - triples
    - (property, object) pairs for a specific subject
    - (subject, property) pairs for specific object
    - etc.
  - the rest is conventional programming…
- Similar tools exist in Java, PHP, etc.

Python example using RDFLib

```
# create a graph from a file
graph = rdflib.Graph()
graph.parse("filename.rdf", format="rdfxml")
# take subject with a known URI
subject = rdflib.URIRef("URI_of_Subject")
# process all properties and objects for this subject
for (s,p,o) in graph.triples((subject,None,None)) :
  do_something(p,o)
```

But programming is not for everyone

- Not everyone wants to program
- On a higher level of abstraction:
  - RDF graphs are “stored”
    - physical triple stores, databases, etc.
    - simple RDF files loaded by underlying tools
    - users can query and update the graph
    - etc.
  - Data files in different formats and storage are “viewed” as RDF
    - “bridges”, APIs between RDB and RDF, CSV and RDF

Example: You publish the raw data…

Data.gov

http://www.data.gov/

Hirek Social Private Mailing lists SW Python RDFa it! Bookmarklets To Mendeley To Faviki bit.ly Shorten with bit.ly MID Save Video

An Official Web Site of the United States Government — Saturday, May 22, 2010 — Text: A+ A- A — Share

DATA.GOV
EMPOWERING PEOPLE

HOME | DATA | TOOLS | COMMUNITY | METRICS | DIALOGUE

LINKING OPEN GOVERNMENT DATA

VIEW MORE

Most Popular Datasets

1. Worldwide M1+ Earthquakes, Past 7 Days
2. U.S. Overseas Loans and Grants (Greenbook)
3. Latest Volumes of Foreign Relations of the...
4. MyPyramid Food Raw Data
5. (SUPERSEDED) 1:2,000,000-scale Hydrologic...

SEARCH OUR CATALOGS

Search our catalogs.. SEARCH

APPS

With so much government data to work with, developers are creating a wide variety of applications, mashups, and visualizations. From crime statistics by neighborhood to the best towns to find a job to seeing the environmental health of your community–these applications arm citizens with the information they need to

COMMUNITY

Data.gov is leading the way in democratizing public sector data and driving innovation. The data is being surfaced from many locations making the Government data stores available to researchers to perform their own analysis. Developers are finding good uses for the datasets, providing interesting and useful applications that allow for new views and public analysis. This is a work in progress, but this movement is spreading to cities, states, and other countries. After just one year a community is born around open government data.

Just look at the numbers:

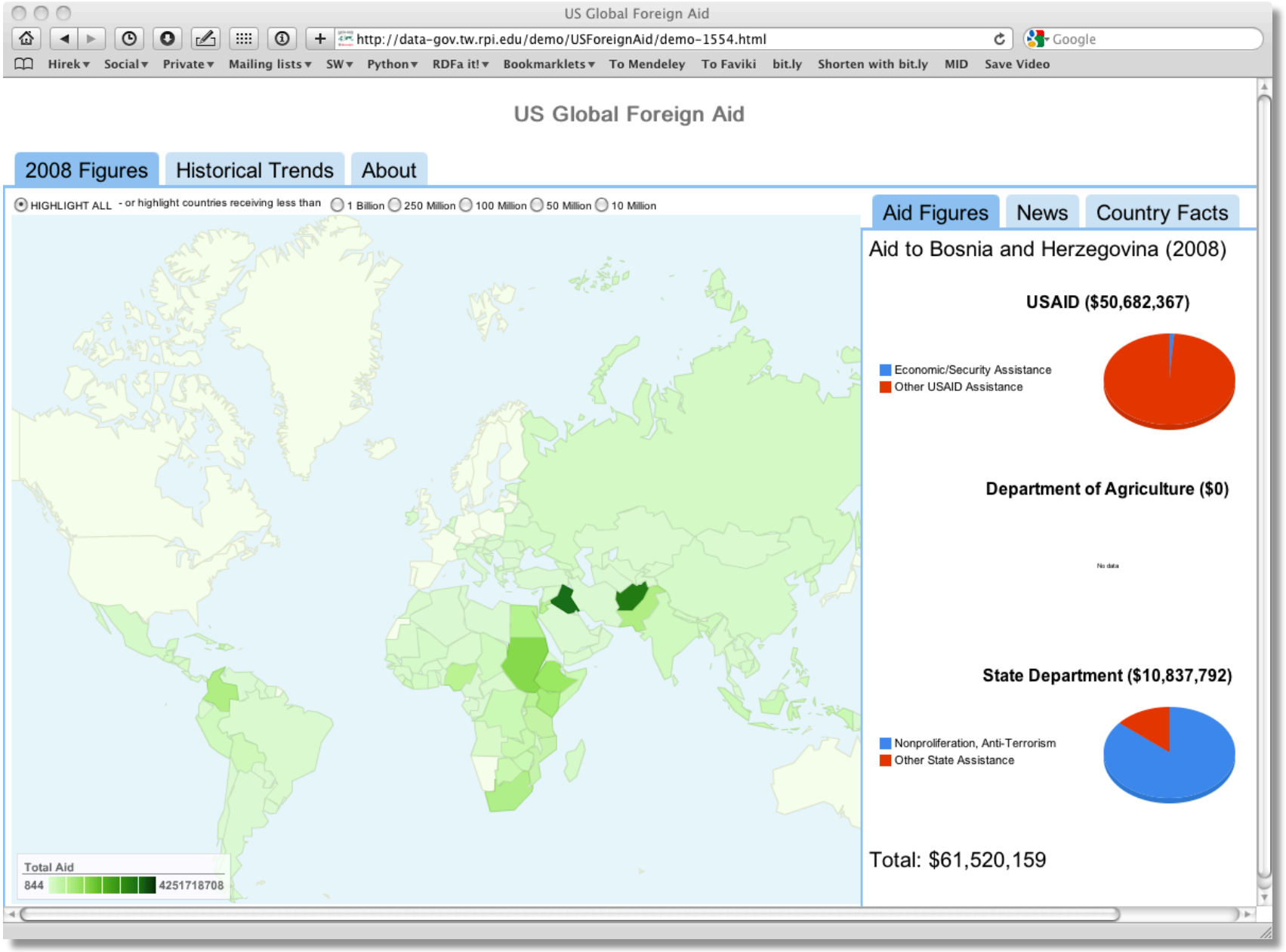
6 Other nations establishing open data
8 States now offering data sites

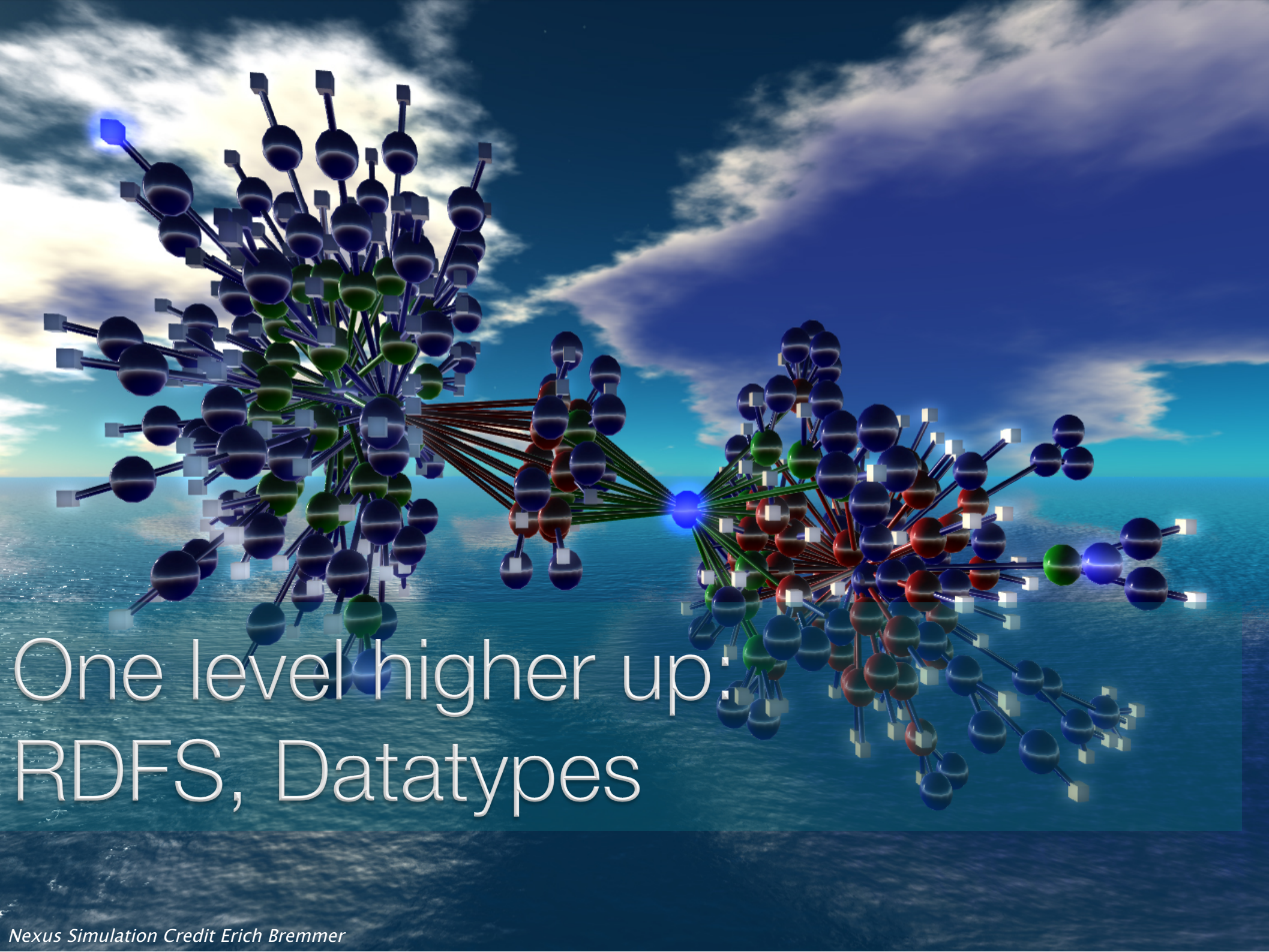
SEMANTIC WEB

As the Web of linked documents evolves to include the Web of linked data, we're working to maximize the potential of Semantic Web technologies to realize the promise of Linked Open Government Data.

Thanks to our collaboration with the **Tetherless World Constellation** at the **Rensselaer Polytechnic Institute**, Data.gov is now hosting one of the largest open collections of RDF datasets in the world! Check out some of their

… and others can use it





One level higher up: RDFS, Datatypes

Nexus Simulation Credit Erich Bremmer

Need for RDF schemas

- First step towards the “extra knowledge”:
  - define the terms we can use
  - what restrictions apply
  - what extra relationships are there?
- Officially: “RDF Vocabulary Description Language”
  - the term “Schema” is retained for historical reasons…

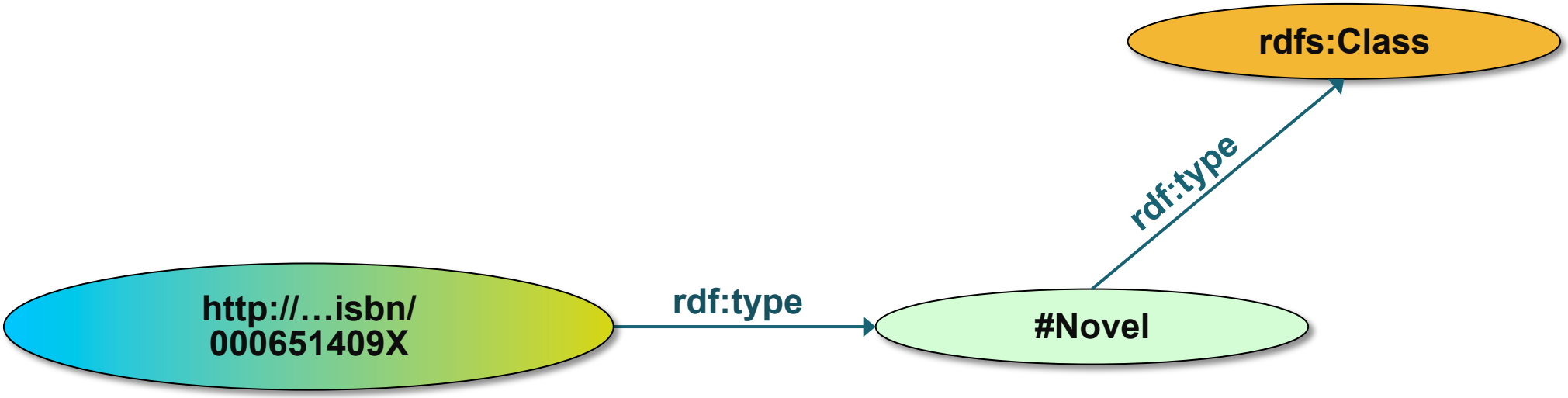
Classes, resources, …

- Think of well known traditional vocabularies:
  - use the term “novel”
  - “every novel is a fiction”
  - “«The Glass Palace» is a novel”
  - etc.
- RDFS defines resources and classes:
  - everything in RDF is a “resource”
  - “classes” are also resources, but…
  - …they are also a collection of possible resources (i.e., “individuals”)
    - “fiction”, “novel”, …

Classes, resources, … (cont.)

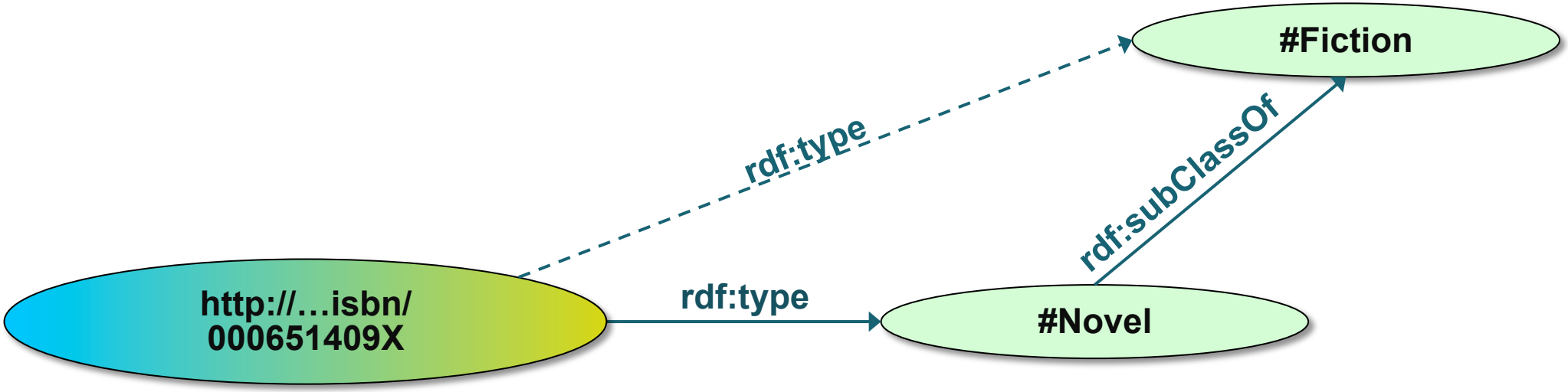
- Relationships are defined among resources:
  - “typing”: an individual belongs to a specific class
    - “«The Glass Palace» is a novel”
    - to be more precise: “«http://.../000651409X» is a novel”
  - “subclassing”: all instances of one are also the instances of the other (“every novel is a fiction”)
- RDFS formalizes these notions in RDF

Classes, resources in RDF(S)



- RDFS defines the meaning of these terms
  - (these are all special URI-s, we just use the namespace abbreviation)

Inferred properties




```
(<http://…/isbn/000651409X> rdf:type #Fiction)
```

- is not in the original RDF data…
- …but can be inferred from the RDFS rules
- RDFS environments return that triple, too

Inference: let us be formal…

- The RDF Semantics document has a list of entailment rules:
  - “if such and such triples are in the graph, add this and this”
  - do that recursively until the graph does not change
- The relevant rule for our example:

```
If:
  uuu rdfs:subClassOf xxx .
  vvv rdf:type uuu .
Then add:
  vvv rdf:type xxx .
```

Properties

- Property is a special class (rdf:Property)
  - properties are also resources identified by URI-s
- There is also a possibility for a “sub-property”
  - all resources bound by the “sub” are also bound by the other
- Range and domain of properties can be specified
  - i.e., what type of resources serve as object and subject

What does this mean?

- Again, new relations can be deduced. Indeed, if

```
:title
  rdf:type     rdf:Property;
  rdfs:domain  :Fiction;
  rdfs:range   rdfs:Literal.

<http://…/isbn/000651409X> :title "The Glass Palace" .
```

then the system can infer that:

```
<http://…/isbn/000651409X> rdf:type :Fiction .
```

Literals

- Literals may have a data type
  - floats, integers, Booleans, etc., defined in XML Schemas
  - full XML fragments
- (Natural) language can also be specified

Examples for datatypes

```
<http://…/isbn/000651409X>
      :page_number "543"^^xsd:integer ;
      :publ_date   "2000"^^xsd:gYear ;
      :price       "6.99"^^xsd:float .
```

Vocabularies

Photo credit "Indy Reading Coalition", Wordpress.com

Vocabularies

- Data integration needs agreements on
  - terms
    - “translator”, “author”
  - categories used
    - “Person”, “literature”
  - relationships among those
    - “an author is also a Person…”, “historical fiction is a narrower term than fiction”
    - ie, new relationships can be deduced

Vocabularies

- There is a need for “languages” to define such vocabularies
  - to define those vocabularies
  - to assign clear “semantics” on how new relationships can be deduced

But what about RDFS?

- Indeed RDFS *is* such framework:
  - there is typing, subtyping
  - properties can be put in a hierarchy
  - datatypes can be defined
- RDFS is enough for many vocabularies

semantic specifications and related DCMI documents [TRANSLATIONS].

References

[DCAM]	http://dublincore.org/documents/2007/06/04/abstract-model/
[HISTORY]	http://dublincore.org/usage/terms/history/
[ISO15836]	http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on
[NAMESPACE]	http://dublincore.org/documents/dcmi-namespace/
[NISOZ3985]	http://www.niso.org/standards/z39-85-2007/
[REVISIONS]	http://dublincore.org/usage/decisions/2012/dcterms-changes/
[RFC5013]	http://www.ietf.org/rfc/rfc5013.txt
[TRANSLATIONS]	http://dublincore.org/resources/translations/

Section 2: Properties in the */terms/* namespace

Term Name:	**abstract**
URI:	http://purl.org/dc/terms/abstract
Label:	Abstract
Definition:	A summary of the resource.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/description
Refines:	http://purl.org/dc/terms/description
Version:	http://dublincore.org/usage/terms/history/#abstract-003
Term Name:	**accessRights**
URI:	http://purl.org/dc/terms/accessRights
Label:	Access Rights
Definition:	Information about who can access the resource or an indication of its security status.
Comment:	Access Rights may include information regarding access or restrictions based on privacy, security, or other policies.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/rights
Refines:	http://purl.org/dc/terms/rights
Has Range:	http://purl.org/dc/terms/RightsStatement
Version:	http://dublincore.org/usage/terms/history/#accessRights-002
Term Name:	**accrualMethod**
URI:	http://purl.org/dc/terms/accrualMethod
Label:	Accrual Method

DCTERMS *is* an RDFS vocabulary

- Each term is an RDFS Property or Class identified by a URI
- Many property terms have range and domain definitions
- Sub-property relationships are also used

```
dcterms:accessRights
    dcterms:issued "2003-02-15"^^xsl:date ;
    a rdf:Property ;
    rdfs:range dcterms:RightsStatement ;
    rdfs:subPropertyOf dcterms:rights ;
    …
```

from http://dublincore.org/2012/06/14/dcterms.ttl

semantic specifications and related DCMI documents [TRANSLATIONS].

References

[DCAM]	http://dublincore.org/documents/2007/06/04/abstract-model/
[HISTORY]	http://dublincore.org/usage/terms/history/
[ISO15836]	http://www.iso.org/iso/search.htm?qt=15836&searchSubmit=Search&sort=rel&type=simple&published=on
[NAMESPACE]	http://dublincore.org/documents/dcmi-namespace/
[NISOZ3985]	http://www.niso.org/standards/z39-85-2007/
[REVISIONS]	http://dublincore.org/usage/decisions/2012/dcterms-changes/
[RFC5013]	http://www.ietf.org/rfc/rfc5013.txt
[TRANSLATIONS]	http://dublincore.org/resources/translations/

Section 2: Properties in the */terms/* namespace

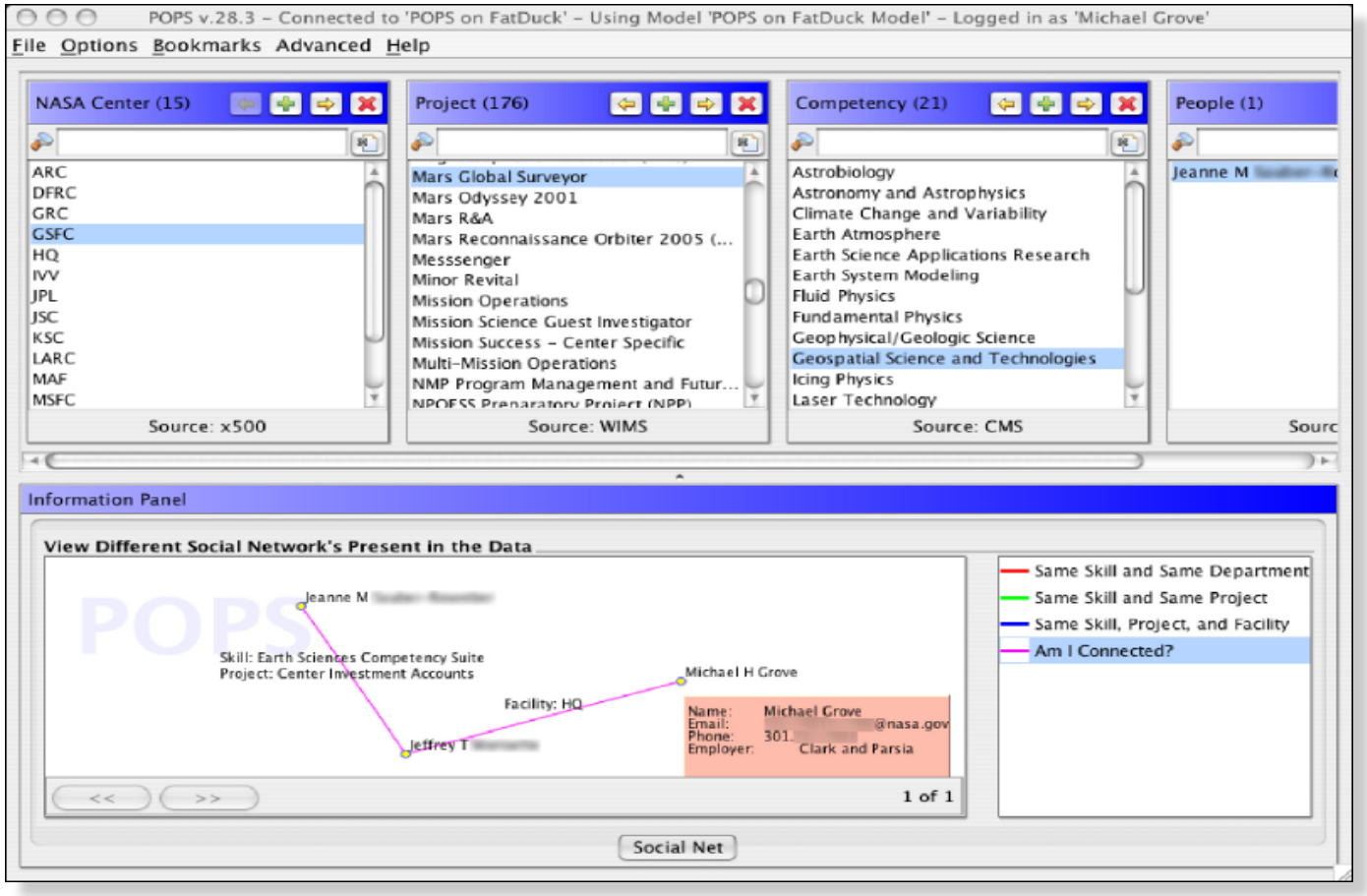
Term Name:	**abstract**
URI:	http://purl.org/dc/terms/abstract
Label:	Abstract
Definition:	A summary of the resource.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/description
Refines:	http://purl.org/dc/terms/description
Version:	http://dublincore.org/usage/terms/history/#abstract-003
Term Name:	**accessRights**
URI:	http://purl.org/dc/terms/accessRights
Label:	Access Rights
Definition:	Information about who can access the resource or an indication of its security status.
Comment:	Access Rights may include information regarding access or restrictions based on privacy, security, or other policies.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/rights
Refines:	http://purl.org/dc/terms/rights
Has Range:	http://purl.org/dc/terms/RightsStatement
Version:	http://dublincore.org/usage/terms/history/#accessRights-002
Term Name:	**accrualMethod**
URI:	http://purl.org/dc/terms/accrualMethod
Label:	Accrual Method

A bit of RDFS can take you far…

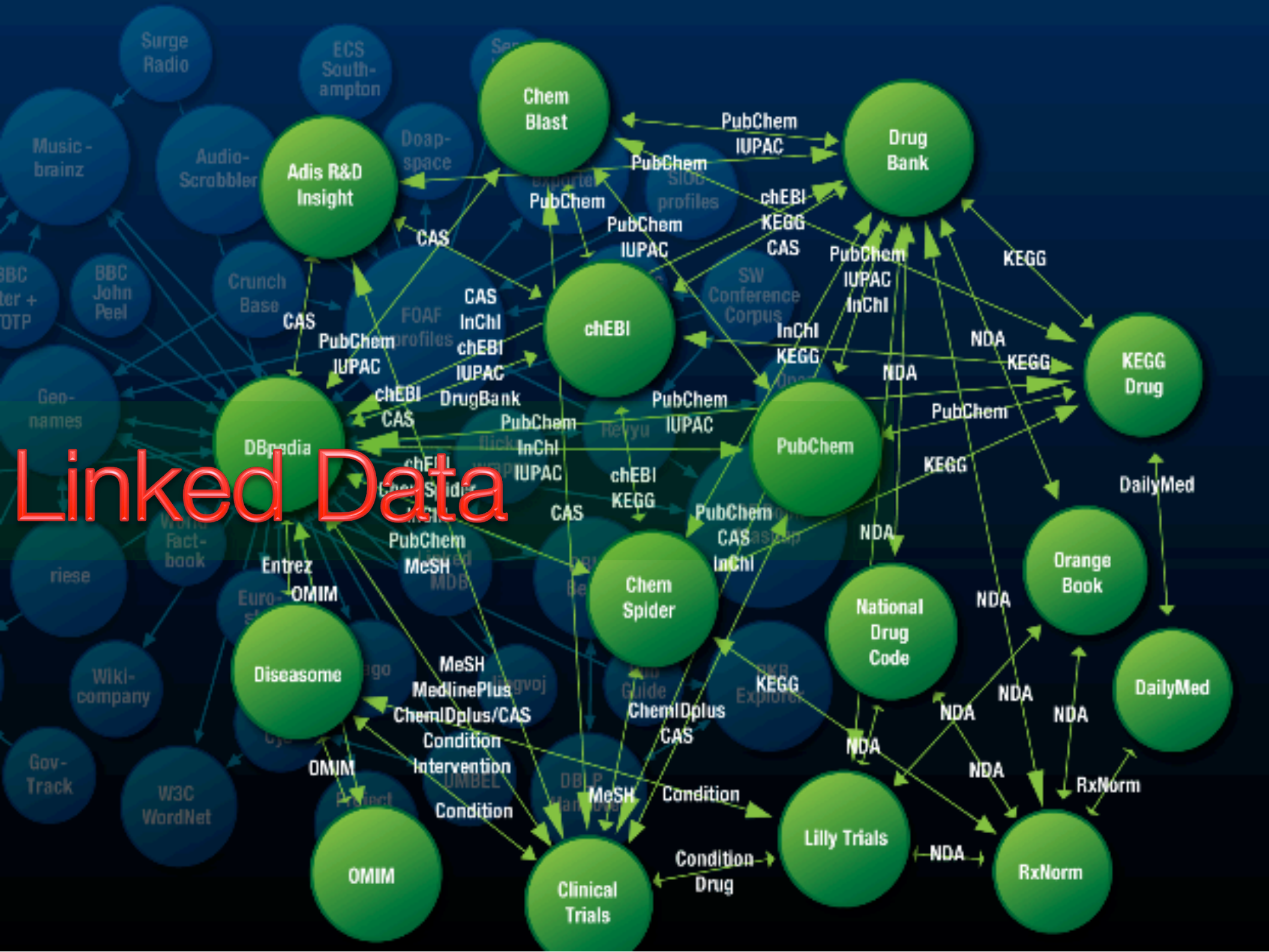
- Remember the power of merge?
- We could have used, in our example:
  - f:auteur is a sub-property of a:author and vice versa (although we will see other ways to do that…)
- Of course, in some cases, more complex knowledge is necessary (see later…)

Example: Find the right experts at NASA

- Expertise locater for nearly 70,000 NASA civil servants, using RDF integration techniques over 6 or 7 geographically distributed databases, data sources, and web services…



(78) *Michael Grove, Clark & Parsia, LLC, and Andrew Schain, NASA, (SWEO Case Study)*

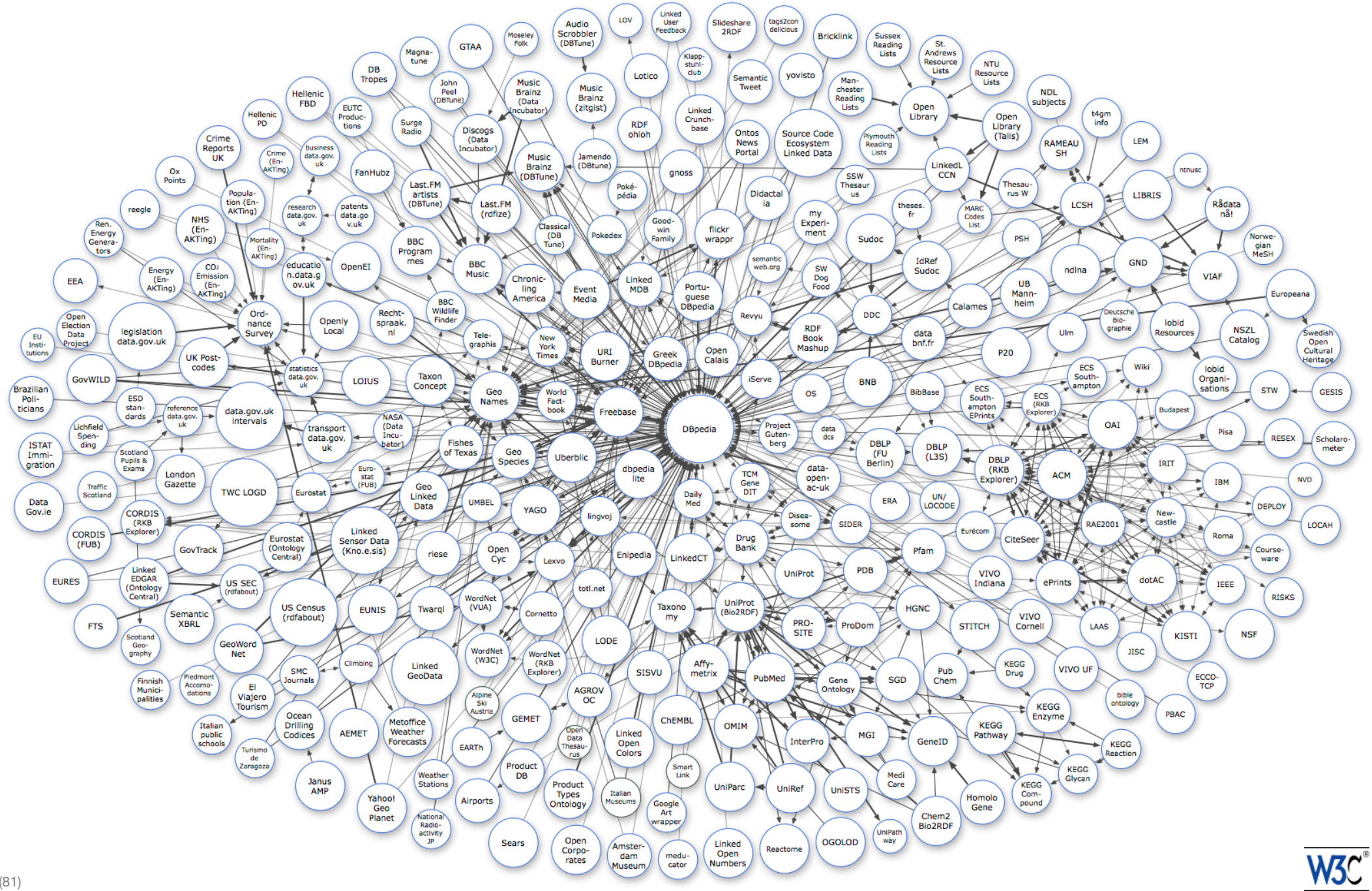

Linked Data
Chem Blast
Drug Bank
Adis R&D Insight
chEBI
KEGG Drug
DBpedia
PubChem
Chem Spider
Orange Book
National Drug Code
DailyMed
Diseasome
Lilly Trials
RxNorm
OMIM
Clinical Trials
PubChem IUPAC
PubChem
chEBI KEGG CAS
CAS
KEGG
PubChem IUPAC InChI
CAS InChI chEBI IUPAC DrugBank
chEBI CAS
PubChem InChI IUPAC
chEBI KEGG
PubChem CAS InChI
PubChem MeSH
Entrez OMIM
MeSH MedlinePlus ChemIDplus/CAS Condition Intervention
ChemIDplus CAS
Condition
Condition Drug
NDA
DailyMed
RxNorm

“Linking Open Data” Project

- Goal: “expose” open datasets in RDF
- Set RDF links among the data items from different datasets

LINKINGOPENDATA
W3C SWEO Community Project

The LOD “cloud”



Linked Data Principles
a.k.a. is your data 5 Star?

★ Available on the Web in some format (i.e., use URI to access the data)

★★ Available as machine-readable structured data (e.g., excel instead of an image scan)

★★★ As before, but using a non-proprietary format (e.g., CSV instead of excel)

★★★★ All the above, plus use open standards (RDF & Co.) to identify things, so that people could point at your stuff

★★★★★ All the above, plus link your data to other people's data to provide context

Linked Data ≈ 5 Star Data
(or a collection of 5 Star Data)

Linked Open Data ≈ Linked Data
without access restrictions

An important distinction

- Linked Data is based on RDF, but
  - in pure RDF, URI-s are used “simply” as unique identifiers
    - it is o.k. to use non-referenceable URI-s like ISBN URN-s, or DOI URN-s
  - *in Linked Data URI-s should really really be de-referenceable*

Example data source: DBpedia

- DBpedia is a community effort to
  - extract structured ("infobox") information from Wikipedia
  - provide a query endpoint to the dataset
  - interlink the DBpedia dataset with other datasets on the Web

Freie Universität Berlin

UNIVERSITÄT LEIPZIG

OPENLINK SOFTWARE

Extracting structured data from Wikipedia

```
@prefix dbpedia <http://dbpedia.org/resource/>.
@prefix dbterm  <http://dbpedia.org/property/>.

dbpedia:Kolkata
  dbpprop:officialName "Kolkata" ;
  dbpprop:longd "88" ;
  dbpprop:longm "22" ;
  dbpprop:longew "E" ;
  foaf:homage <http://www.kmcgov.in> ;
  dbpedia-owl:populationTotal "4486679" ;
  dbpprop:areaTotalKm "185" ;
  ...
dbpedia:Amitav_Ghosh
  dbbedia-owl:birthPlace dbpedia:Kolkata ;
  ...
```

Kolkata
কলকাতা
Calcutta

— City —

Country	India
State	West Bengal
Division	Presidency
District	Kolkata[A]
Government	
· Type	Mayor–Council
· Body	KMC
· Mayor	Sovan Chatterjee[1]
· Sheriff	Indrajit Ray[2]
· Police commissioner	Surajit Kar Purakayastha[3]
Area	
· City	185 km² (71 sq mi)
· Metro	1,886.67 km² (728.45 sq mi)
Elevation	9 m (30 ft)
Population (2011)[4]	
· City	4,486,679
· Rank	7th
· Density	24,000/km² (63,000/sq mi)
· Metro[5]	14,112,536
· Metro rank	3rd
· Metropolitan	14,617,882 (3rd)
Demonym	Calcuttan
Time zone	IST (UTC+05:30)
ZIP code(s)	7000 xx, 7001 xx
Area code(s)	+91-33
Vehicle registration	WB 01–79
UN/LOCODE	IN CCU
Official language	Bengali and English
Major Ethnic Settlements	Bengali, Marwari, Bihari and Others
Website	www.kmcgov.in

Links among open datasets

```
<http://dbpedia.org/resource/Kolkata>
  owl:sameAs freebase:Kolkata ;
  owl:sameAs <http://sws.geonames.org/1275004> ;
  owl:sameAs <http://ru.dbpedia.org/resources/Калькутта> ;
  ...
```

```
<http://sws.geonames.org/1275004>
  rdfs:seeAlso <http://dbpedia.org/resource/Kolkata>
  wgs84_pos:lat "22.56263" ;
  wgs84_pos:long "88.36304";
  gn:parentCountry <http://sws.geonames.org/1269750> ;
 ...
```

Processors can switch automatically from one to the other…

The importance of Linked Data

- It provides a core set of data that applications can build on
  - stable references for “things”,
    - e.g., http://dbpedia.org/resource/Kolkata/
  - many many relationships that applications may reuse
    - e.g., the BBC application!
  - a “nucleus” for a larger, semantically enabled Web!

Back to the BBC…

BBC – Music – Eric Clapton

www.bbc.co.uk/music/artists/618b6900-0618-4f1e-b835-bccb17f84294

helped reggae reach a mass market. Two of his most popular recordings were "Layla", recorded by Derek and the Dominos, another band he formed, and Robert Johnson's "Crossroads", recorded by Cream. Following the death of his son Conor in 1991, Clapton's grief was expressed in the song "Tears in Heaven", which featured in his Unplugged album.

Read more at Wikipedia...

WIKIPEDIA This entry is from Wikipedia, the user-contributed encyclopedia. It may not have been reviewed by professional editors and is licensed under the GNU Free Documentation License. If you find the biography content factually incorrect, defamatory or highly offensive you can edit this article at Wikipedia. Find out more about our use of this data.

Links & Information

LINKS

Official homepage at ericclapton.com

Fanpage at whereseric.com

Youtube at youtube.com/user/ericclapton

Twitter at twitter.com/EricClaptonNews

MySpace at myspace.com/ericclapton

Wikipedia article on Eric Clapton

Last.fm page on Eric Clapton

Discogs at discogs.com/artist/Eric Clapton

MusicBrainz entry on Eric Clapton

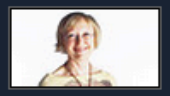
MEMBER OF Cream (1966-1968), The Yardbirds (1963-1965), Derek and the Dominos, John Mayall & The Bluesbreakers (1965-1966), Blind Faith (1968-1969), The Louisiana Gator Boys

COLLABORATED ON The Dirty Mac (1968)

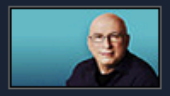
Richard Allinson
BBC Radio 2

Gerry Anderson
BBC Radio Ulster

The Chris Evans Breakfast Show
BBC Radio 2



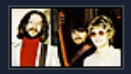
The Late Show with Cherrie McIlwaine
BBC Radio Ulster



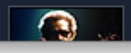
Ken Bruce
BBC Radio 2

Information displayed about artists played on BBC programmes is incomplete at present. Find out more about this artist play count information.

Similar Artists



Derek and the Dominos



J.J. Cale

Back to the BBC…

About: Eric Clapton

dbpedia.org/page/Eric_Clapton

Private webmail | History | LocalData | TR | 2013 | TR List | Private | Social | Mailing lists | SW | Python | RDFa it! | » | Other Bookmarks

About: Eric Clapton

An Entity of Type : artist, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

DBpedia

Eric Patrick Clapton, CBE, (born 30 March 1945) is an English guitarist and singer-songwriter. Clapton is the only three-time inductee to the Rock and Roll Hall of Fame: once as a solo artist, and separately as a member of The Yardbirds and Cream. Clapton has been referred to as one of the most important and influential guitarists of all time. Clapton ranked second in Rolling Stone magazine's list of the "100 Greatest Guitarists of All Time" and fourth in Gibson's Top 50 Guitarists of All Time.

Property	Value
dbpedia-owl:abstract	▪ Eric Patrick Clapton, is een Britse gitarist, componist en zanger van blues-, rock- en popmuziek. ▪ Eric Clapton,, est un guitariste, chanteur et compositeur de blues et de rock britannique, né le 30 mars 1945 à Ripley, près de Londres. En 2003, le mensuel Rolling Stone l'a classé quatrième meilleur guitariste de tous les temps. Ses fans le surnomment « God » (Dieu). ▪ Eric Patrick Clapton CBE, přezdívaný Slowhand (česky pomalá ruka) je britský kytarista, zpěvák a hudební skladatel. Je považován za jednoho z nejlepších světových bluesových kytaristů. ▪ Eric Patrick Clapton, CBE ist ein englischer Blues- und Rock-Gitarrist und -Sänger. Er ist mehrfacher Grammygewinner und dreifaches Mitglied der „Rock and Roll Hall of Fame". Clapton prägte die Entwicklung des Bluesrocks seit den 1960er Jahren wesentlich mit und gilt als einer der bedeutendsten Gitarristen. Auf der im Jahr 2011 aktualisierten Liste der 100 Greatest Guitarists of All Time der amerikanischen Musikzeitschrift Rolling Stone findet sich Clapton auf Rang 2. ▪ Eric Patrick Clapton er en engelsk gitarist, sanger og låtskriver. ▪ Eric Patrick Clapton, pseudonim Slowhand – brytyjski muzyk rockowy i bluesowy, gitarzysta, kompozytor, wokalista. Członek wielu zespołów rockowych i jeden z najbardziej znanych symboli muzyki rockowej. W 2003 pojawił się na 4. miejscu listy 100 największych gitarzystów wszech czasów magazynu "Rolling Stone", zaś w 2011 na miejscu 2.. W 2000 został wprowadzony jako artysta solowy do Rock and Roll Hall of Fame. ▪ Soprannominato Slowhand (Mano lenta) e God (Dio), da Chuck Berry definito The Man Of The Blues (L'uomo del Blues), Eric Clapton è uno dei più famosi e influenti chitarristi blues. Plurivincitore di Grammy Award, è l'unico musicista che vanta 3 inserimenti nella Rock'n'Roll Hall of Fame. La rivista Rolling Stone lo ha inserito al numero #2 nella lista Rolling Stone's 100 Greatest Guitarists of All Time, che annovera i migliori chitarristi di tutti i tempi. ▪ エリック・パトリック・クラプトン（Eric Patrick Clapton, CBE、1945年3月30日 - ）はイギリスのミュージシャン、ギタリスト。 ▪ Э́рик Па́трик Клэ́птон CBE — британский рок-музыкант. Командор Ордена Британской империи. Его прозвище — «медленная рука». Клэптон стал одним из самых уважаемых и влиятельных музыкальных деятелей в эру рока. Его стиль со временем менялся, но он всегда держался блюзовых корней. Играл с блюз-роковыми командами, с хард-роковыми, много работал сольно и как сессионный музыкант. ▪ 에릭 클랩튼은 영국의 기타리스트이자 싱어송라이터로 역사상 가장 영향력 있는 기타리스트 중 한 사람으로 꼽힌다. 로큰롤 명예의 전당의 유일한 3중 헌액자이다(솔로 아티스트 자격, 야드버즈의 멤버 자격, 크림의 멤버 자격) . 그는 롤링 스톤지가 선정한 100인의 가장 위대한 기타리스트 4위, 100인의 위대한 아티스트 53위에 오르기도 했다. 클랩튼의 음악 스타일은 시간이 지남에 따라 여러 가지로 바뀌었지만, 대개 블루스에서 출발한다. 그럼에도 불구하고 다양한 장르에서 혁신적으로 활동했다. 그는 존 메이올 & 블루스브레이커즈, 야드버즈와 함께 블루스 록의 창시자이며, 크림과 함께 하드 록의 창시자로 꼽히기도 한다. 또한 그는 델타 블루스와 사이키델릭 록(〈Sunshine of Your Love〉), 팝(〈Change the World〉), 레게(〈I Shot the Sheriff〉) 등으로도 큰 성공을 거뒀다. ▪ 埃里克·帕特里克·克莱普顿，CBE（英語：Eric Patrick Clapton，1945年3月30日－），英国吉他手、歌手和作曲家。他曾经获得过格莱美奖，是20世纪最成功的音乐家之一，在摇滚名人堂裡有三项成就。他是有史以来最伟大的吉他手之一，2003年在滚石杂志评选的一百大吉他手位列第四，在百大摇滚艺术家里面排名53位。虽然克莱普顿的曲风多变，但是他的大部分作品还是基于布鲁斯风格。他是一个音乐的改革者，他的主要贡献包括布鲁斯摇滚（在「约翰·梅奥尔和布鲁斯破坏者乐队」期间）以及迷幻摇滚（在「奶油乐队」期间）。他还成功地从Delta布鲁斯（代表作「我和约翰先生」）音乐风格转向流行音乐（代表作「改变世界」）和雷鬼乐（代表作「我射杀警长」）。 ▪ Eric Patrick Clapton, CBE, (born 30 March 1945) is an English guitarist and singer-songwriter. Clapton is the only three-time inductee to the Rock and Roll

Back to the BBC…

dbpedia.org/data/Eric_Clapton.n3

```
                yago:Person100007846 ,
                yago:LivingPeople ,
                yago:EnglishGuitarists ,
                dbpedia-owl:MusicalArtist ,
                dbpedia-owl:Person ,
                dbpedia-owl:Artist ,
                dbpedia-owl:Agent .
@prefix ns13:   <http://umbel.org/umbel/rc/> .
dbpedia:Eric_Clapton    rdf:type        ns13:Artist ,
                yago:CommandersOfTheOrderOfTheBritishEmpire ,
                <http://dbpedia.org/class/yago/2000sSingers> ,
                <http://dbpedia.org/class/yago/1980sSingers> ,
                foaf:Person ,
                <http://dbpedia.org/class/yago/2010sSingers> .
@prefix ns14:   <http://schema.org/> .
dbpedia:Eric_Clapton    rdf:type        ns14:MusicGroup ,
                <http://dbpedia.org/class/yago/1990sSingers> ,
                <http://dbpedia.org/class/yago/PeopleFromGuildford(district)> ,
                yago:AlumniOfKingstonUniversity ,
                yago:EnglishPeopleOfCanadianDescent ,
                ns13:MusicalPerformer ,
                owl:Thing ,
                ns14:Person ,
                yago:Artist109812338 ,
                yago:EnglishRockSingers ,
                yago:PeopleAssociatedWithTheBeatles ,
                yago:EnglishBluEsSingers ,
                yago:EnglishBluEsGuitarists .
@prefix ns15:   <http://ja.dbpedia.org/resource/> .
dbpedia:Eric_Clapton    owl:sameAs      ns15:エリック・クラプトン ,

<http://ru.dbpedia.org/resource/\u041A\u043B\u044D\u043F\u0442\u043E\u043D,_\u042D\u0440\u0438\u043A> ,
                <http://rdf.freebase.com/ns/m.02qwg> .
@prefix ns16:   <http://sw.cyc.com/concept/> .
```

Same dataset, another site

Eric Clapton – seevl.net

http://seevl.net/entity/QwznaqLQ

TR 2011 Read it later▾ To...▾ Social▾ Private▾ Mailing lists▾ Python▾ SW▾ RDFa it!▾ Bookmarklets▾ Wolfram

seevl
REINVENTING MUSIC DISCOVERY

about / blog / FAQ / jobs / dev zone

Semantic search

Feedback

Eric Clapton

Like Tweet 0 Share 0

Eric Patrick Clapton, CBE (born 30 March 1945) is an English guitarist, vocalist, and songwriter. Clapton is the only three-time inductee to the Rock and Roll Hall of Fame: once as a solo artist, ...

read more

Videos

powered by YouTube

Fact sheet

Birth(s):	1945-03-30
Birth place(s):	England Ripley, Surrey

Related bands and artists

George Harrison

Show the relations !

George Harrison, MBE (25 February 1943 – 29 November 2001) was an English rock guitarist, singer-songwriter, actor and film producer who achieved international fame as lead guitarist of The Beatles. Often referred to as "the quiet Beatle",Schaffner, The Boys from Liverpool, pp. 77-78.</ref> Following the band's break-up, he ...

more info

John Lennon

Show the relations !

John Winston Ono Lennon, MBE (9 October 1940 – 8 December 1980) was an English musician and singer-songwriter who rose to worldwide fame as one of the founding members of The Beatles, one of the most commercially successful and critically acclaimed acts in the history of popular music. Along with fellow ...

more info

Jack Bruce

Show the relations !

John Symon Asher "Jack" Bruce (born 14 May 1943, Bishopbriggs, Scotland) is a Scottish musician and songwriter, respected as a founding member of the British psychedelic rock power trio, Cream, for a solo career that spans several decades, and for his participation in several well-known musical ensembles. Best recognized as ...

Loading "http://seevl.net/entity/QwznaqLQ", completed 28 of 29 items

Same dataset, another site

Eric Clapton – seevl.net

http://seevl.net/entity/QwznaqLQ

TR 2011 Read it later ▾ To... ▾ Social ▾ Private ▾ Mailing lists ▾ Python ▾ SW ▾ RDFa it! ▾ Bookmarklets ▾ Wolfram

seevl
REINVENTING MUSIC DISCOVERY

about / blog / FAQ / jobs / dev zone

Semantic search

Feedback

Eric Clapton

Like Tweet 0 Share 0

Eric Patrick Clapton, CBE (born 30 March 1945) is an English guitarist, vocalist, and songwriter. Clapton is the only three-time inductee to the Rock and Roll Hall of Fame: once as a solo artist, ...

read more

Videos

powered by YouTube

Fact sheet

Birth(s):	1945-03-30
Birth place(s):	England Ripley, Surrey

Related bands and artists

George Harrison

Show the relations !

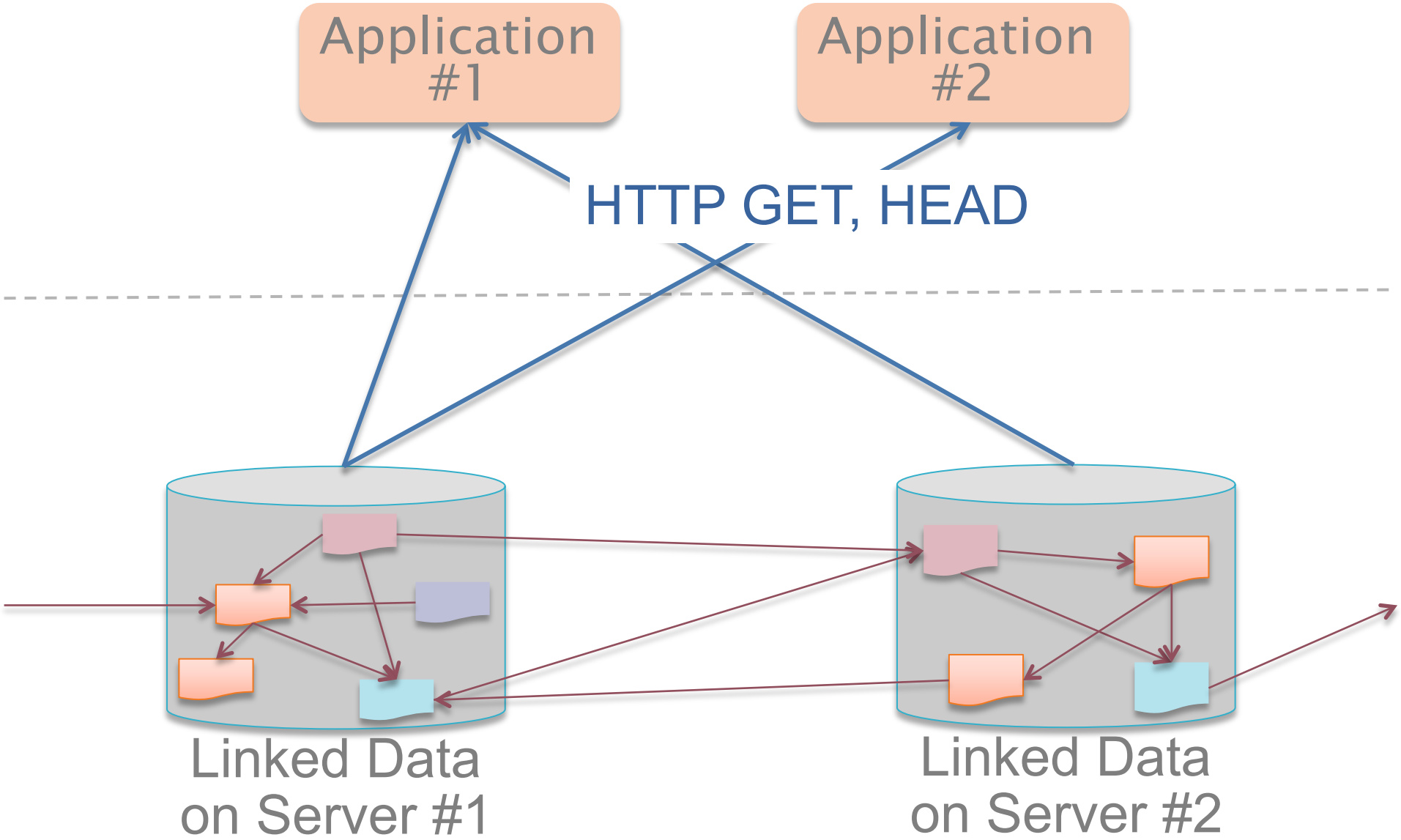
- Both have been signed on the same **label**: Apple Records
- Both **collaborated with**: Jim Gordon
- Both have similar **topic**:
  - 1980s singers
  - 1960s singers
  - Grammy Award winners
  - Slide guitarists
  - 1970s singers
  - English rock guitarists
  - 2000s singers
  - Rock and Roll Hall of Fame inductees
  - 1990s singers
  - English singer-songwriters
  - English male singers
  - Lead guitarists
- Both play the same **genre**:
  - Singing
  - Rock music
  - Psychedelic rock
- Both play the same **instrument**:
  - Guitar
  - Fender Stratocaster
  - Slide guitar
  - Gibson Les Paul

Loading "http://seevl.net/entity/QwznaqLQ", completed 26 of 27 items

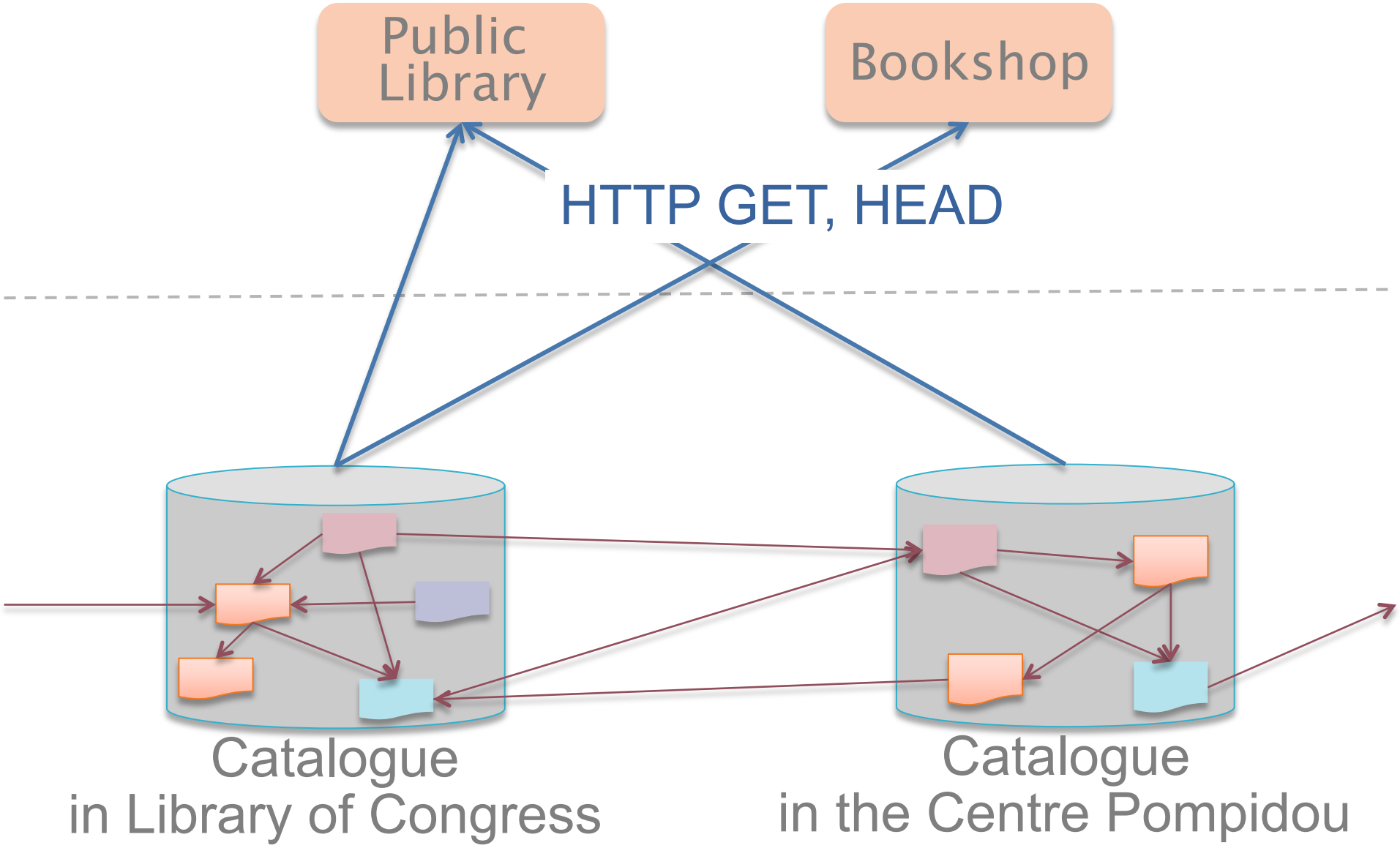
Simplest way to use Linked Data

- Use Web technologies! E.g., use HTTP…
  - extract the data via HTTP GET
  - interpret the (hopefully RDF) content to, e.g., extract new relationships and resources
- Tools are available to do that
  - although, truth must be said, it is still an evolving world

Example: Simple Application Integration via Linked Data



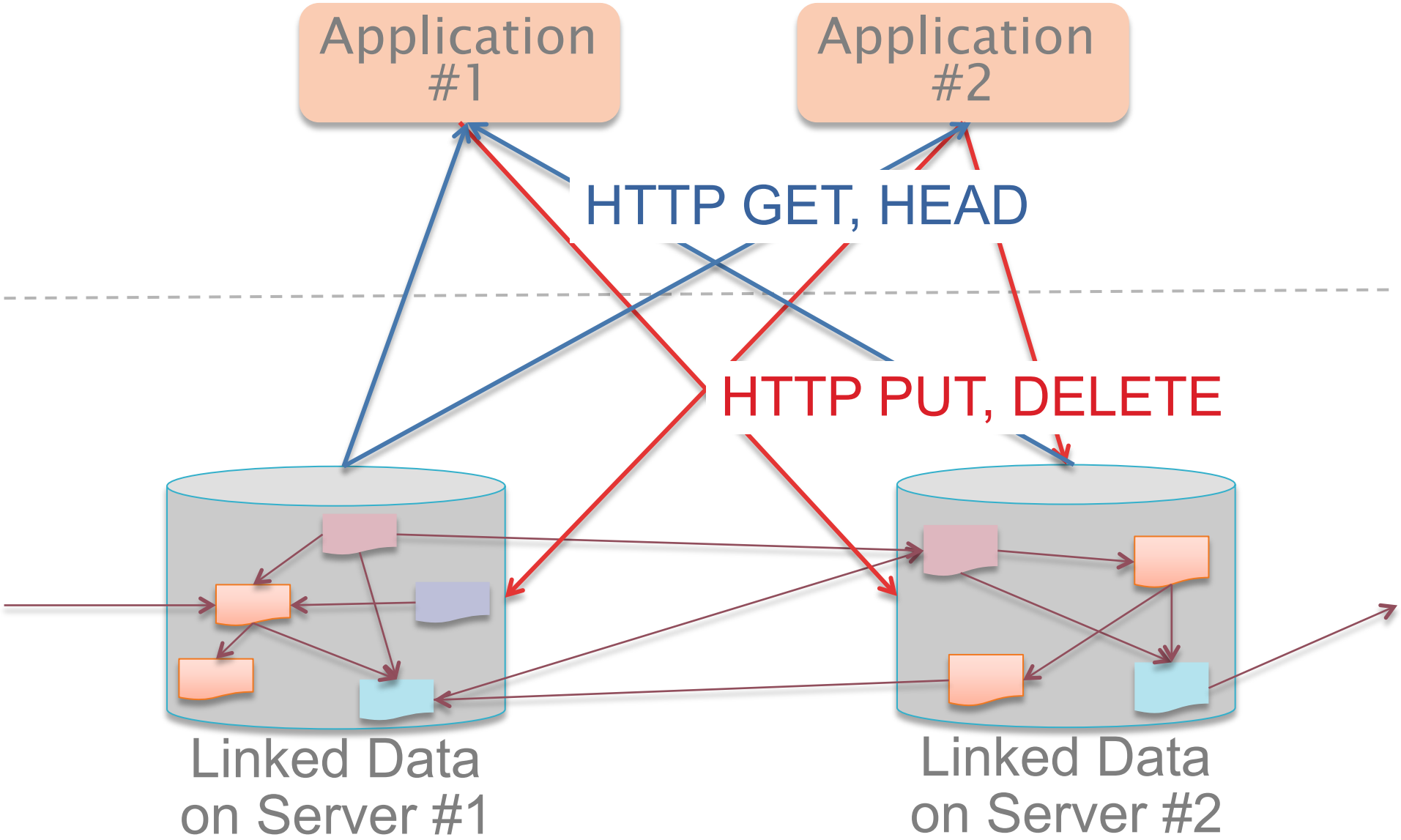
Example: Simple Application Integration via Linked Data



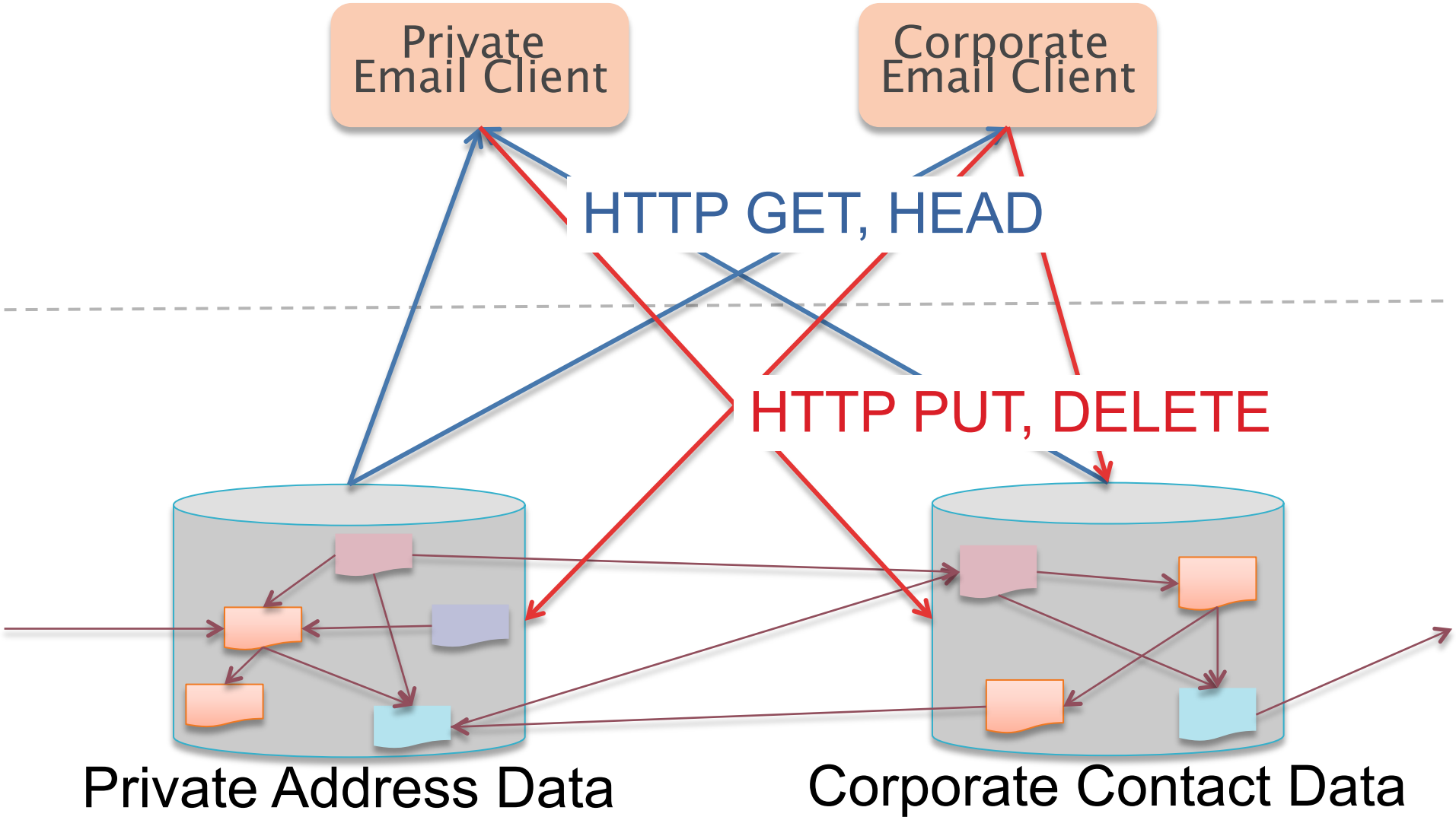
Read only vs. Read/Write

- Currently, Linked Data is dominated by *publishing* data for read-only usage
  - creating/updating the data is done “out of band”
- The future to read *and* write Linked Data

Example: Simple Application Integration via Linked Data



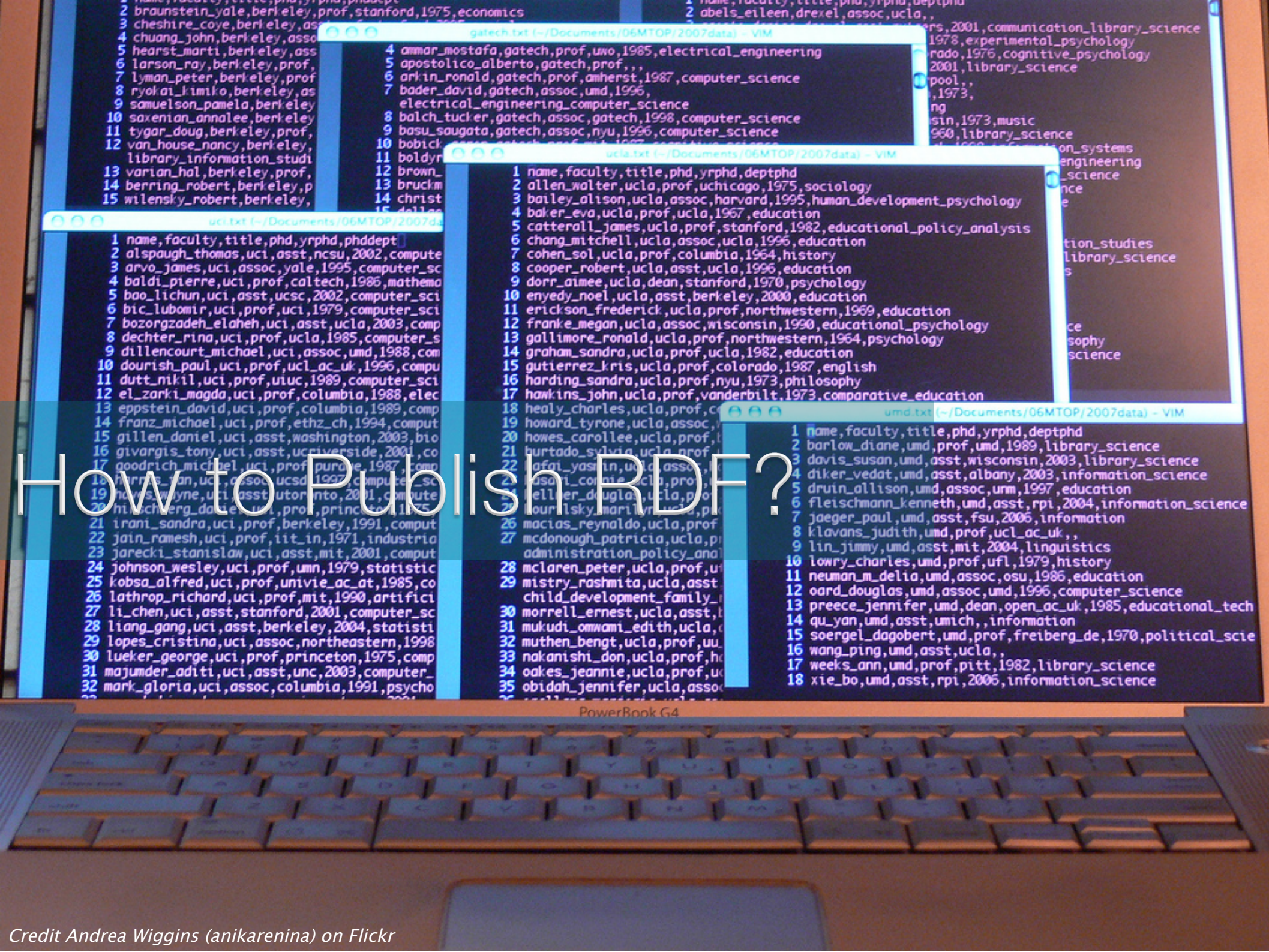
Example: Simple Application Integration via Linked Data



Linked Data Platform (LDP)

- Define an HTTP/RESTful based infrastructure to publish, read, write, or modify linked data
  - typical usage: data intensive application in a browser, application integration using shared data…
- The infrastructure should be easy to implement and install
  - provides an “entry point” for Linked Data applications!
- The work is still ongoing at W3C…

How to Publish RDF?



Credit Andrea Wiggins (anikarenina) on Flickr

Technically: the simple approach

- Write RDF/XML, Turtle, etc. “manually”
  - in some cases that is necessary, but it really does not scale…

Access to Relational Databases

Photo credit "mayhem", Flickr

Relational Databases and RDF

- Many of the data on the Web is, in fact, in RDB-s
- Proven technology, huge systems, many vendors…
- Data integration on the Web must provide access to RDB-s
  - RDB data should be “exported” into, e.g., RDF

What is “export”?

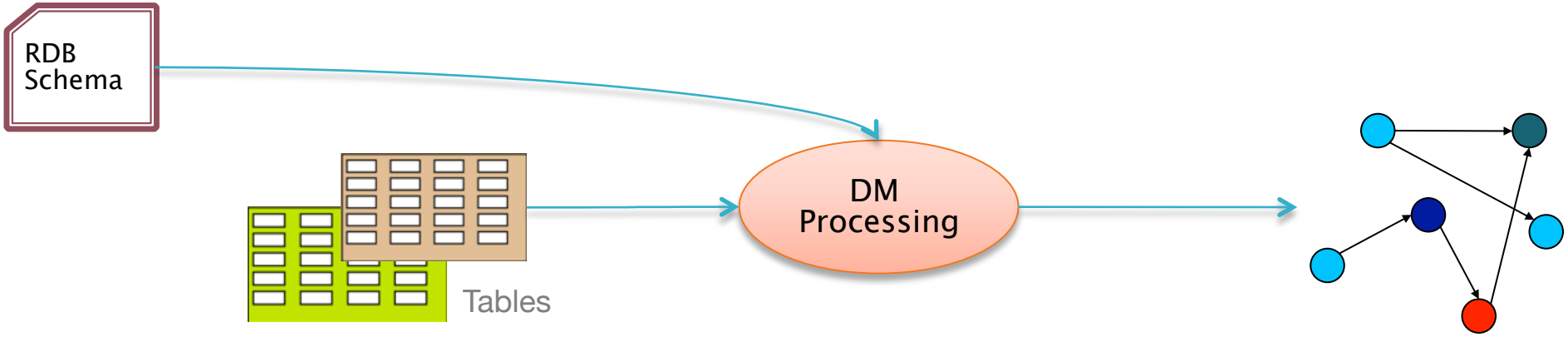
- “Export” does not *necessarily* mean physical conversion
  - for very large databases a “duplication” would not be an option
  - systems may provide “bridges” to make RDF queries on the fly
  - result of export is a “logical” view of the RDB content

Simple export: RDF Direct Mapping

- A standard RDF “view” of RDB tables
- Valid for all RDB-s, independently of the RDB schema
- Fundamental approach:
  - each row is turned into a series of triples with a common subject (subject URI based on primary key value)
  - column names provide the predicate names
  - cell contents are the objects as literals
  - cross-referenced tables are expressed through URI subjects

What DM processor does

- An DM processor has access to:
  - an RDB schema
  - a database governed by the schema
- … and produces an RDF graph using a standard mapping



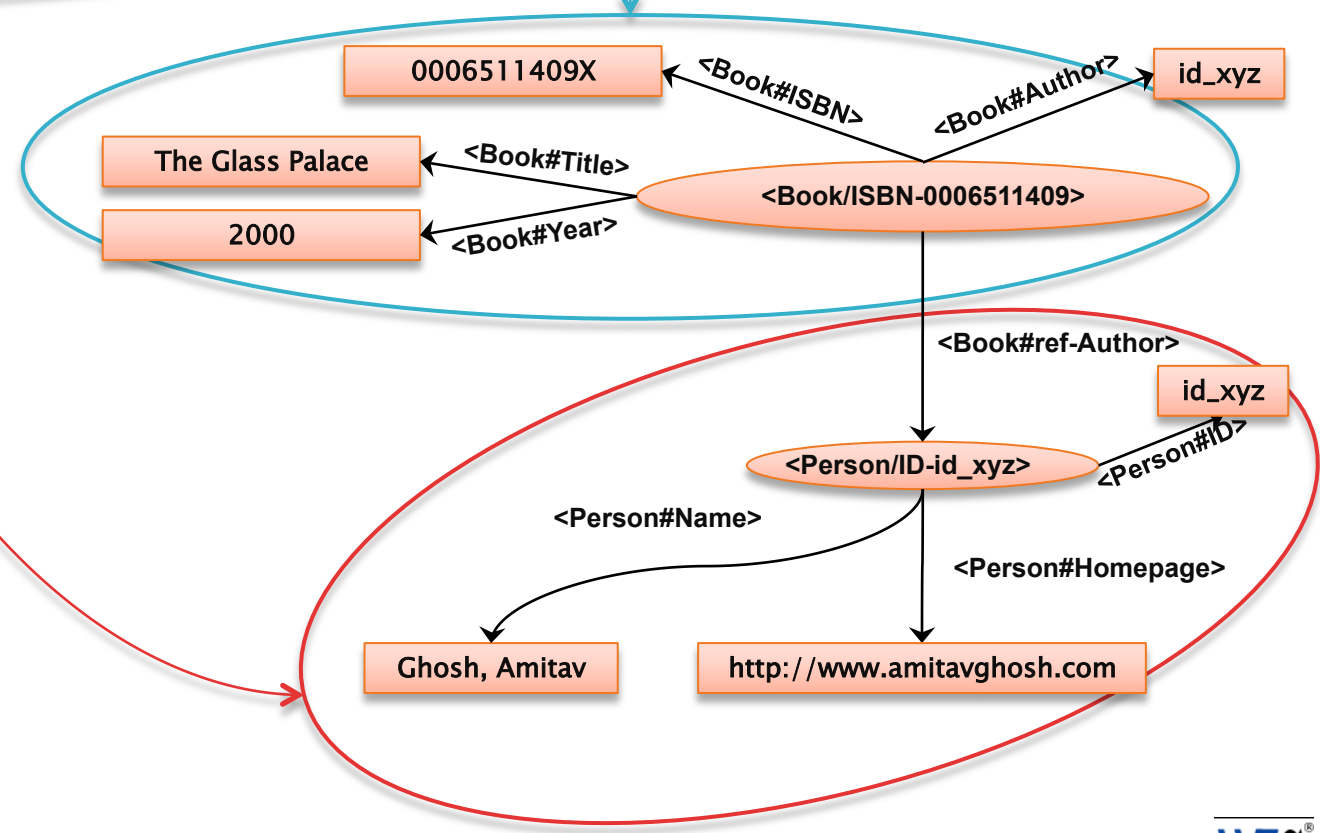
Result of the Direct Mapping

- What do we get?
  - we have an RDF “view” of the RDB tables
  - *a query against the RDF view may be transformed into an SQL query against the original tables*

Direct mapping of the bookshop tables

ISBN	Author	Title	Publisher	Year
0006511409X	id_xyz	The Glass Palace	id_qpr	2000

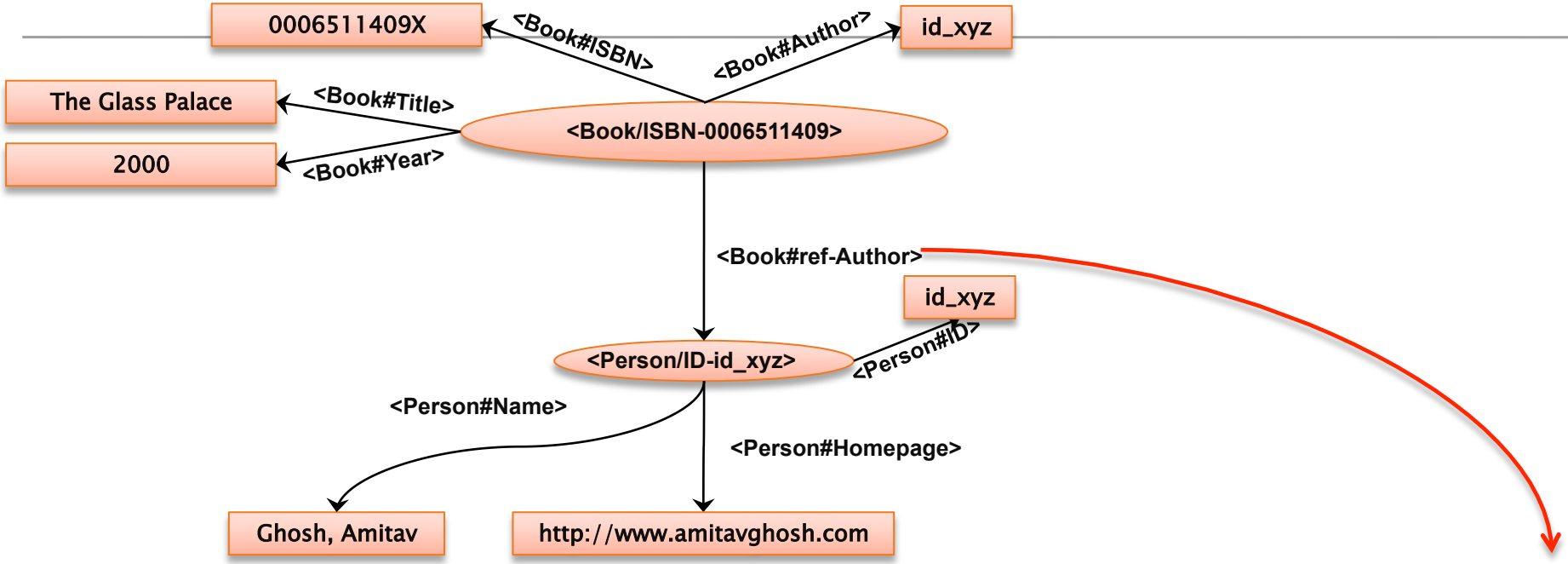
ID	Name	Homepage
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com



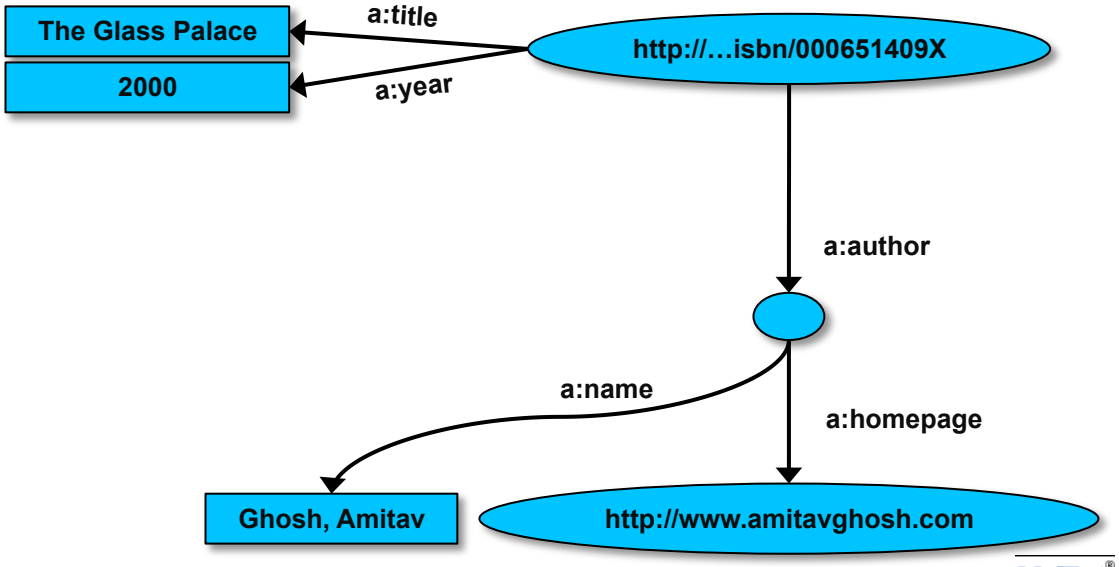
Result of the Direct Mapping

- What do we miss?
  - an RDF view that is close to our application; a more “natural” view of the data
  - i.e., *the result of the Direct Mapping must be transformed, somehow, into an RDF that an application may use*

Direct graph must be transformed



- Property names should be mapped
- URI-s should be minted
- Literals should be replaced by URI-s



Enters R2RML

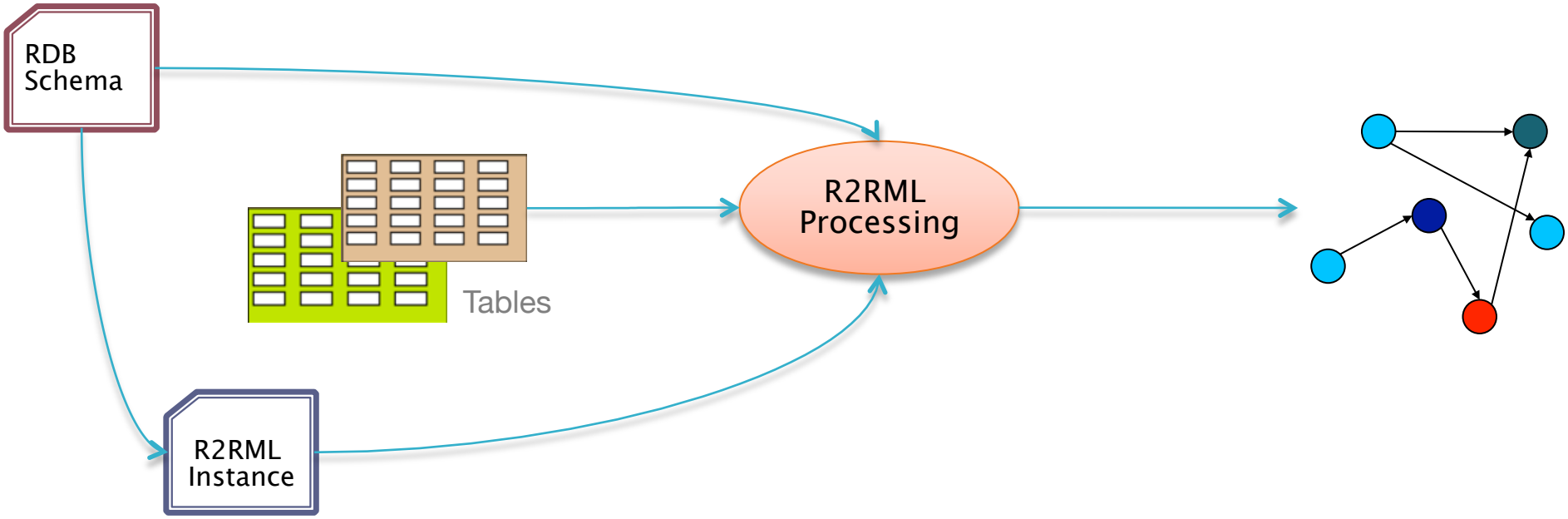
- Separate vocabulary for a finer control of the mapping
  - gets to the final RDF graph with one processing step
- Fundamentals are similar:
  - each row is turned into a series of triples with a common subject
  - cross-referenced tables linked via URI-s

Enters R2RML

- There is a finer control over the structure of the result graph
  - the format of the (common) subject URI can be controlled
  - objects might be URI-s generated on the fly via templates from column names
  - datatypes can be assigned to literal objects
  - “virtual” tables can be generated through SQL before processing them through R2RML
- R2RML can generate the *final* RDF ready to be used by an application

What R2RML processor does

- An R2RML processor has access to:
  - an RDB schema
  - an R2RML instance
  - a database governed by the schema
- … and produces an RDF graph



CSV Data

- CSV ("Comma Separated Value") may be the single most widespread format of datasets published on the Web
- There is no standard RDF mapping yet, but…
- … the structure is very close to a relational table
- W3C plans to start a group that will (also) standardize that mapping
  - pretty much along the lines of the Direct Mapping

Structured data in HTML: RDFa & microdata

Photo credit "shetladd", Flickr

RDF with HTML

- By adding some “meta” information, the same *source* can be reused
  - typical example: your personal information, like address, should be readable for humans and processable by machines
- Some solutions have emerged:
  - add extra statements in *microdata* or *RDFa* that can be converted to RDF
    - microdata can be used for a (useful) subset of RDF
    - RDFa is, essentially, a complete serialization of RDF

Ivan Herman

Who am I?

I graduated as mathematician at the Eötvös Loránd University of Budapest, Hungary, in 1979. After a brief scholarship at the Université Paris VI I joined the Hungarian research institute in computer science (SZTAKI) where I worked for 6 years (and turned into a computer scientist...). I left Hungary in 1986 and, after a few years in industry in Munich, Germany, I joined the Centre Mathematics and Computer Sciences (CWI) in Amsterdam where I have a tenure position since 1988. I received a PhD degree in Computer Science in 1990 at the University of Leiden, in the Netherlands. I joined the World Wide Web Consortium (W3C) Team as Head of W3C Offices in January 2001 while maintaining my position at CWI. I served as Head of Offices until June 2006, when I was asked to take the Semantic Web Activity Lead position, which is now my principal work at W3C.

Before joining W3C I worked in quite different areas (distributed and dataflow programming, language design, system programming), but I spend most of my research years in computer graphics and information visualization. I also participated in various graphics related ISO standardization activities and software developments. My "professional" home page contains a list of my publications (see also my Mendeley account), my public presentations, and details of the various projects I participated in the past. There is also a dblp entry for my publications generated automatically (although I am not sure it is complete...). (B.t.w., based on my publications, my Erdős number is ≤4...)

In my previous life (i.e., before joining W3C...) I was member of the Executive Committee of the Eurographics Association for 15 years, and I was vice-chair of the Association between 2000 and 2002. I was the co-chair of the 9th World Wide Web Conference, in Amsterdam, May 2000; since then, I have also been member of IW3C2 (International World Wide Web Conference Committee), responsible for the World Wide Web Conference series. Since autumn 2007 I am also member of SWSA (Semantic Web Science Association), the committee responsible for the International Semantic Web Conferences (better known as "ISWC") series.

Some personal data

- The Hungarian spelling of my full name is Herman Iván. Ie, my name is Ivan (well, spelled properly: Iván) and my surname is Herman (many in the Netherlands and in Germany mix it up, and use "Herman" as my name... this is aggravated by the fact that, uniquely in Europe, the Hungarian custom is to put surname first).
- Nationalities: French and Hungarian
- Gender: male
- Family: I am married and have a son, David.
- Date and city of birth: 24th February, 1955, Budapest, Hungary
- Email addresses: 'ivan' on my own ivan-herman.net domain, 'ivan' on the w3.org domain, or 'ivan.herman' on the cwi.nl domain
- (Mobile) Phone: +31-641044153
- Skype ID: ivan_herman
- I live in Amstelveen (see also geonames), the Netherlands (lat: 52.302063, long: 4.87397). This is a suburb of Amsterdam. The closest airport is Amsterdam Schiphol
- I am the administrator of the Semantic Web Activity Blog at W3C which can either be accessed directly or via its

some links

- personal homepage
- more data on me
- personal blog (RSS feed)
- homepage at W3C
- "professional" homepage
- "official" CV
- more about me
- my photos

"social" links

- facebook
- flickr
- picasa web
- linkedin
- IWIW
- tripit
- twitter
- Mendeley
- Google+
- freebase

Ivan Herman

www.ivan-herman.net/foaf-rdfa-1.1/foaf.html

Delicious | LocalData | TR | 2012 | My Mercurial | Private | Mailing lists | Social | SW | Python | RDFa it! | Bookmarklets

Ivan Herman

Ivan Herman

Who am I?

I graduated as mathematician at the Eötvös Loránd University of Budapest, Hungary, in 1979. Aft… ef scholarship at the Université Paris VI I joined the Hungarian research institute in computer science (SZTAKI … worked for 6 years (and turned into a computer scientist…). I left Hungary in 1986 and, after a few years … Munich, Germany, I joined the Centre Mathematics and Computer Sciences (CWI) in Amsterdam wh… position since 1988. I received a PhD degree in Computer Science in 1990 at the University of Leiden … joined the World Wide Web Consortium (W3C) Team as Head of W3C Offices in January … sition at CWI. I served as Head of Offices until June 2006, when I was asked to take … position, which is now my principal work at W3C.

Before joining W3C I worked in quite … ming, language design, system programming), but I spend … nformation visualization. I also participated in various … opments. My "professional" home page contai… entations, and details of the various … ted automatically (although I am not

… mmittee of the Eurographics Association … as the co-chair of the 9th World Wide Web … of IW3C2 (International World Wide Web … es. Since autumn 2007 I am also member of … the International Semantic Web Conferences

… name is Ivan (well, spelled properly: Iván) and my … mix it up, and use "Herman" as my name… this is … custom is to put surname first).

… 'ivan' on the w3.org domain, or 'ivan.herman' on the … 3.92663, long: 4.87397). This is a suburb of … can either be accessed directly or via its

```
<div
    <h1
    <meta p
    <h2>Who am
    <p>I graduated
    I joined the <a rel
        <span property="dc:
    </a> Team as Head of <a re
    <link rel="owl:sameAs" href="
    <link rel="owl:sameAs" href="http
    <link rel="foaf:workplaceHomepage"
    <meta property="schema:jobTitle" conte

    <p>Before joining W3C I worked in quite di
    </p>

    <p>In my previous life (i.e., before joining W3C…)
    </p>
    <h2>
    Some personal data
    </h2>
    <ul>
        <li>The Hungarian spelling of my full name is <span prop
        Ie, my name is <span property="foaf:givenname schema:given
        <span property="foaf:givenname schema:givenName" lang="hu">
        <span property="foaf:surname schema:familyName">Herman</span
        (many in the Netherlands and in Germany mix it up, and use "He
        </li>
```

some links

- personal homepage
- more data on me
- personal blog (RSS feed)
- homepage at W3C
- "professional" homepage
- "official" CV
- more about me
- my photos

"social" links

- facebook
- flickr
- picasa web
- linkedin
- IWIW
- tripit
- twitter
- Mendeley
- Google+
- freebase

Source of: http://www.ivan-herman.net/foaf-rdfa-1.1/foaf.html

```
                <meta property="foaf:accountName" content="ivan-herman" />
            </a>
        </li>
        <li>
            <a href="https://plus.google.com/u/0/113268051484517627727" typeof="foaf:OnlineAccount">
                <span property="foaf:accountServiceHomepage" href="http://www.mendeley.com/">Google+</span>
                <meta property="foaf:accountName" content="113268051484517627727" />
            </a>
        </li>
        <li>
            <a about="http://www.ivan-herman.net/foaf#me" rel="owl:sameAs" resource="http://rdf.freebase.com/ns/en.
        </li>
    </ul>
   </div>
</div>

<div id="content" >
   <h1 property="schema:name foaf:name">Ivan Herman</h1>
   <meta property="foaf:title" content="Dr" />
   <h2>Who am I?</h2>
   <p>I graduated as mathematician at the <a rel="foaf:schoolHomepage schema:alumniOf" href="http://www.elte.hu/"><spa
   I joined the <a rel="schema:worksFor" href="http://www.w3.org" resource="http://www.w3.org/Data#W3C">
      <span property="dc:title">World Wide Web Consortium (W3C)</span>
   </a> Team as Head of <a rel="foaf:pastProject" href="http://www.w3.org/Consortium/Offices"><span property="dc:title"
   <link rel="owl:sameAs" href="http://www.ivan-herman.net/me" />
   <link rel="owl:sameAs" href="http://www.ivan-herman.net/Ivan_Herman" />
   <link rel="foaf:workplaceHomepage" href="http://www.w3.org"/>
   <meta property="schema:jobTitle" content="Semantic Web Activity Lead" />

   <p>Before joining W3C I worked in quite different areas (distributed and dataflow programming, language design, syst
   </p>

   <p>In my previous life (i.e., before joining W3C…) I was member of the Executive Committee of the <a rel="foaf:pastI
   </p>
   <h2>
   Some personal data
   </h2>
   <ul>
       <li>The Hungarian spelling of my full name is <span property="foaf:name" lang="hu">Herman Iván</span>.
       Ie, my name is <span property="foaf:givenname schema:givenName">Ivan</span> (well, spelled properly:
       <span property="foaf:givenname schema:givenName" lang="hu">Iván</span>) and my surname is
       <span property="foaf:surname schema:familyName">Herman</span>
       (many in the Netherlands and in Germany mix it up, and use “Herman” as my name… this is aggravated by the fact t
       </li>
```

Yielding…

```
<http://www.ivan-herman.net/foaf#me>
    schema:alumniOf     <http://www.elte.hu> ;
    foaf:schoolHomePage <http://www.elte.hu> ;
    schema:worksFor     <http://www.w3.org/W3C#data> ;
    …
<http://www.elte.hu>
    dc:title "Eötvös Loránd University of Budapest" .
…
<http://www.w3.org/W3C#data>
    dc:title "World Wide Web Consortium (W3C)”
…
```

The Telegraph

Oscars 2012: The Artist, review

The Artist, an utterly beguiling silent, black-and-white celebration of early Hollywood won Best Picture at the Oscars 2012.

★★★★★

Source of http:

Oscars 2012: The Artist, review – Telegraph

```
                        <li class
                        <li><a hr

                <div class="cl">
        </div>
</div>

<!-- googleon: all -->
<div id="tmglBody" >
        <div class="access"><a name="a

        <div class="twoThirdsThird2 gutt
                <div class="twoThirds gut
                        <div class="storyh

        <h1 itemprop="name">Oscars 2012: The
                <h2 itemprop="description">
The Artist, an utterly beguiling silent, black
  Hollywood won Best Picture at the Oscars 2012

</h2>
        <div class="rating" itemprop="reviewRating
                        <meta itemprop="worstRating
                        <meta itemprop="bestRating"
                        <span itemprop="ratingValue"
                        <img src="/template/verl-0/i/
                </div>
        <div class="artIntro">
                                <div id="storyEm
                        <div class="slideshow ssIntro">
                                <div class="nextPrevLayer"
                                                <di

                                                <
                                                </div>
                                <div>
                <div>

        </div>
                        </div>
<div class="oneHalf gutter">
                                <div class="story">
                                        <div class="cl"> </div>
                <!-- remove the whitespace added by escenic before end of
```

http://www.telegraph.co.uk/culture/film/filmreviews/8982558/Oscars-2012-The-Artist-r

the artist

2012 | My Mercurial | Private | Mailing lists | Social | SW | Python | RDFa it! | Bookmarklets | To... | Web Data Inspector

The Telegraph

Search - enhanced by Google

Monday 09 April 2012

BLOGS | CULTURE | TRAVEL | LIFESTYLE | FASHION | TECH | Dating | Offers | Jobs

Theatre | Hay Festival | Dance | Opera | Photography | Comedy | Video | In the Know

Coming Soon | Talking Movies | Interviews | DVDs | Film Life | Film Video

The Artist, review

black and white celebration of early Hollywood won Best Picture at the Oscars

TELEGRAPH TICKETS »

Source of http://www.telegraph.co.uk/culture/film/filmreviews/8982558/Oscars-2012-The-Artist-review.html

```
                        <li class="first"><a href="/">Home</a><span>&raquo;</span></li>
                        <li><a href="http://www.telegraph.co.uk/culture/">Culture</a><span>&raquo;</span></li>
                                        <li><a href="http://www.telegraph.co.uk/culture/film/">Film</a><span>&raquo;</span></li>
                                        <li class="styleSix"><a href="http://www.telegraph.co.uk/culture/film/filmreviews/">Film R
                <div class="cl"></div>
        </div>
</div>

<!-- googleon: all -->
<div id="tmglBody" >
        <div class="access"><a name="article"></a></div>

        <div class="twoThirdsThird2 gutterUnder">
                <div class="twoThirds gutter"  itemscope itemtype="http://schema.org/Review">
                        <div class="storyHead">

        <h1 itemprop="name">Oscars 2012: The Artist, review</h1>
                <h2 itemprop="description">
The Artist, an utterly beguiling silent, black-and-white celebration of early
  Hollywood won Best Picture at the Oscars 2012.

</h2>
        <div class="rating" itemprop="reviewRating" itemscope itemtype="http://schema.org/Rating">
                        <meta itemprop="worstRating" content = "0.5">
                        <meta itemprop="bestRating" content = "5">
                        <span itemprop="ratingValue" class="hidden">5</span>
                        <img src="/template/verl-0/i/ratings/star_5_styleSix.png" alt="5 out of 5 stars" width="73" height="14" />
                </div>
        <div class="artIntro">
                                        <div id="storyEmbSlide">
                        <div class="slideshow ssIntro">
                                <div class="nextPrevLayer">
                                                <div class="ssImg">
                                                                <img src="http://i.telegraph.co.uk/multimedia/archive/0209
                                                                <div class="artImageExtras" >

                                                                        <div class="ingCaptionCredit">
                                                                                <span class="caption">Bérénice Bejo as ris
                                                                </div>
                                                        </div>
                                                </div>
                                        </div>
                <div>

        </div>
                        </div>
<div class="oneHalf gutter">
                                <div class="story">
                                        <div class="cl"> </div>
                <!-- remove the whitespace added by escenic before end of </a> tag -->
```

Yielding…

```
[ rdf:type schema:Review ;
  schema:name "Oscars 2012: The Artist, review" ;
  schema:description "The Artist, an utterly beguiling…" ;
  schema:ratingValue "5" ;
  …
]
```

schema.org

- Schema.org is a cooperation of search engines (Bing, Google, Yahoo!, and Yandex)
- *It is a large vocabulary* that they all understand
- The terms are extracted from HTML5+microdata or HTML5+RDFa
  - the various partners use it for different purposes
  - *it can be used by anyone outside of the search world!*

Review – schema.org

schema.org/Review

schema.org

Search

Home Schemas Documentation

Thing > CreativeWork > Review

A review of an item – for example, a restaurant, movie, or store.

Property	Expected Type	Description
Properties from Thing		
additionalType	URL	An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax – the 'typeof' attribute – for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
sameAs	URL	URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Freebase page, or official website.
url	URL	URL of the item.
Properties from CreativeWork		
about	Thing	The subject matter of the content.
accountablePerson	Person	Specifies the Person that is legally accountable for the CreativeWork.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
alternativeHeadline	Text	A secondary title of the CreativeWork.
associatedMedia	MediaObject	The media objects that encode this creative work. This property is a synonym for encodings.
audience	Audience	The intended audience of the item, i.e. the group for whom the item was created.
audio	AudioObject	An embedded audio object.
author	Organization or Person	The author of this content. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.
award	Text	An award won by this person or for this creative work.
awards	Text	Awards won by this person or for this creative work. (legacy spelling; see singular form, award)
[illegible]	CreativeWork or	A citation or reference to another creative work, such as another publication, web page, scholarly article, etc.

the artist movie - Google Search

https://www.google.nl/#hl=en&sugexp=frgbld&gs_nf=1&cp=11&gs_id=5p&xhr=t&q=the

movie datatabase

Delicious | LocalData | TR | 2012 | My Mercurial | Private | Mailing lists | Social | SW | Python | RDFa it! | Bookmarklets | To... | Web Data Inspector

Everything
Images
Maps
Videos
News
Shopping
More

Amsterdam
Change location

Any time
Past hour
Past 24 hours
Past week
Past month
Past year
Custom range...

More search tools

The Artist showtimes for Amsterdam

Pathe Tuschinski - Reguliersbreestraat 26-34, Amsterdam - Map
11:50 - 14:05 - 19:10

Filmtheater "De Uitkijk" - Prinsengracht 452, Amsterdam - Map
12:15 - 19:00 - 21:15

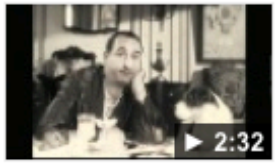
Filmtheater Rialto - Ceintuurbaan 338, Amsterdam - Map
12:45

+ Show more theaters

The Artist (2011) - IMDb
www.imdb.com/title/tt1655442/
Silent **movie** star George Valentin bemoans the coming era of talking ... Still of Jean Dujardin and Missi Pyle in **The Artist** Still of Bérénice Bejo in **The Artist** Reem ...
Full cast and crew - The Artist Trailer (Official ... - Bérénice Bejo - Jean Dujardin

The Artist (**film**) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/**The_Artist_(film)**
The Artist is a 2011 French romantic comedy drama in the style of a black-and-white silent **film** written and directed by Michel Hazanavicius, starring Jean ...
Jean Dujardin - Bérénice Bejo - Uggie - Diegesis

The Artist Trailer 2011 HD - YouTube



www.youtube.com/watch?v=O8K9AZcSQJE
25 Aug 2011 - 3 min - Uploaded by TrailersApplecom
I love how George Clooney, and Brad Pitt, lost the Best actor catagory to this **film**. It just shows that there is ...

More videos for **the artist movie** »

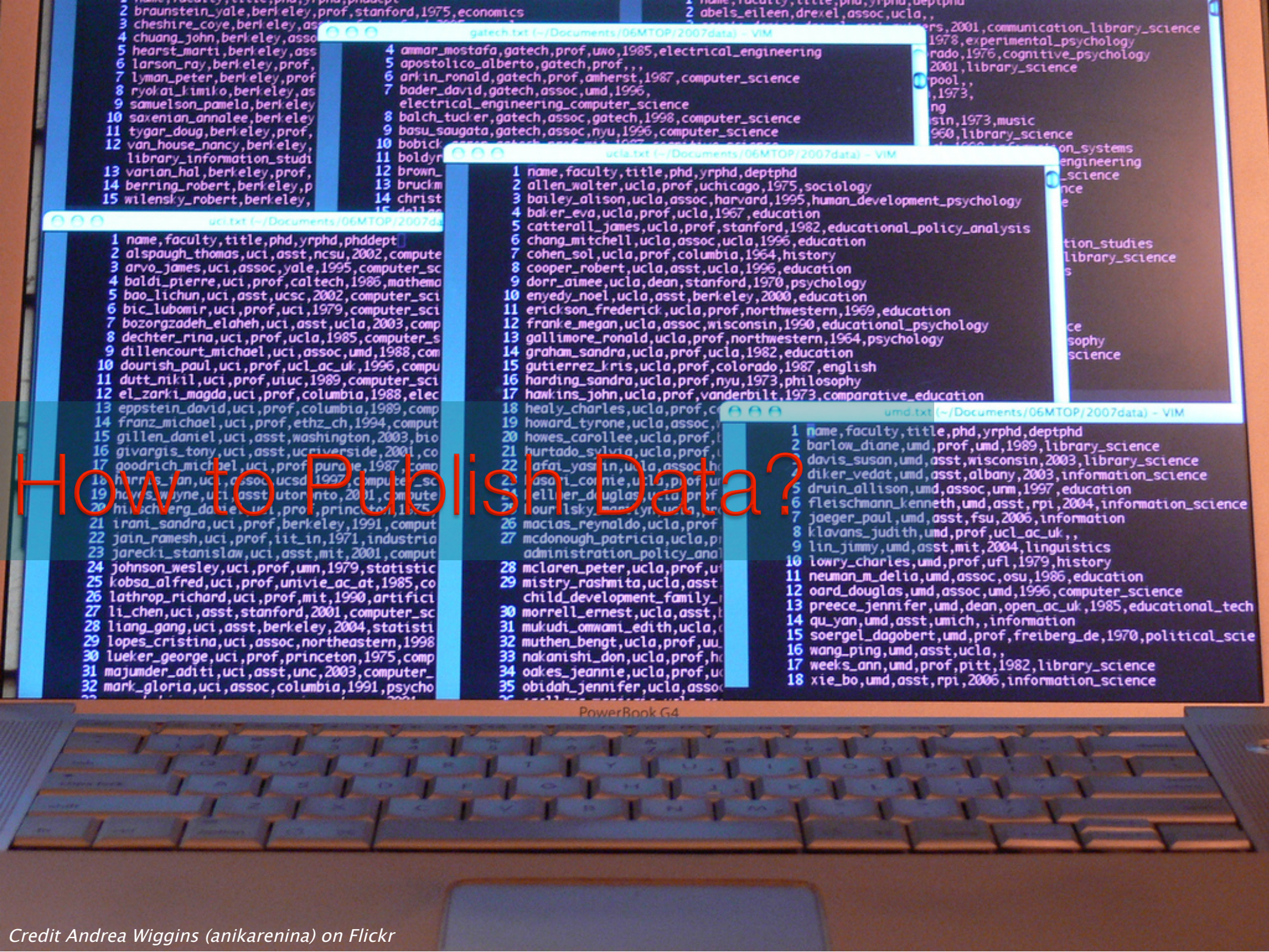
Oscars 2012: **The Artist**, review - Telegraph
www.telegraph.co.uk › Culture › Film › Film Reviews
★★★★★ Review by Robbie Collin
27 Feb 2012 – **The Artist**, the final **film** to be released in 2011 and also the most heart-swellingly joyful one, is a silent **movie**, screened in black and white and ...

The Artist is the perfect **film** about Hollywood | Hadley Freeman

HTML+* can also be generated

- CMS systems may generate such data automatically
  - e.g., Drupal 7 generates pages with RDFa included
- There are a number of plugins to blogging systems
  - generate HTML+RDFa or HTML+microdata

How to Publish Data?



Credit Andrea Wiggins (anikarenina) on Flickr

Some things to remember if you publish data

- Publish your data first, care about sexy user interfaces later!
  - the “raw data” can become useful on its own right and others may use it
  - you can add your added value later by providing nice user access
- If possible, publish your data in RDF but if you cannot, others may help you in conversions
  - trust the community…
- *Add links to other data*. “Just” publishing isn’t enough…

Some things to remember if you publish data

- Think about persistence and versioning
  - others may depend on the data you publish…
- Be careful about URIs you choose
  - avoid being very domain dependent; e.g., your institution's name may change at some point
  - use some consistent URI structure
    - e.g. http://publications.europa.edu/{type}/{subtype}
  - etc.
- Avoid reinventing the wheel when choosing vocabularies

Some things to remember if you publish data

- *Document your data*, i.e., provide metadata
  - there are vocabularies to do this
    - Data Catalog Vocabulary (DCAT)
    - Vocabulary of Interlinked Datasets (VoID)
    - DCTERMS (of course ☺)
    - vocabularies for licensing (Open Data Commons, government licenses)
      - this area is still very much in development…
- A new group on Best Practices will start at W3C soon

ODI Open Data Certificate

https://certificates.theodi.org/certificates/31

Private webmail | 2013 | 2012 | LocalData | TR | TR List | Private | Social | Mailing lists | SW | Python | RDFa it! | Bookmarklets | Other Bookmarks

open data certificate™ BETA

Register | Login

Create new certificate | About | Get in touch

Embed this on your site

Pilot level
self certified
GB

Northern Ireland Hospital Waiting Lists

This data has achieved Pilot level on 11 June 2013 which means extra effort went in to support and encourage feedback from people who use this open data.

General Information

This data is described at
http://data.gov.uk/dataset/northern_ireland_waiting_lists

This data is curated by
Department of Health, Social Services and Public Safety

The data curator's website is
http://www.dhsspsni.gov.uk/

Legal Information

This data was
originally created or generated by its curator

Copyright and database rights are described at
http://data.gov.uk/dataset/northern_ireland_waiting_lists

This data is available under
UK Open Government Licence

The content is available under
UK Open Government Licence

This data contains
aggregated data

aggregated data

Practical Information

The data appears in this collection
http://data.gov.uk/data/search?q=waiting+list&tags=hospital

The lag between creation and publication of this data is
minimal

Data quality issues are documented at
http://www.dhsspsni.gov.uk/index/stats_research/hospital-stats/...

Quality control processes are described at
http://www.dhsspsni.gov.uk/index/stats_research/hospital-stats/...

The data is available
for a long time

Technical Information

Releases follow this consistent URL pattern
http://www.dhsspsni.gov.uk/northern_ireland_waiting_times_qe_...

Releases of this data are listed at
http://www.dhsspsni.gov.uk/index/stats_research/hospital-stats/...

This data is
machine-readable

Statistical data is published
in a presentation format

Social Information

Send questions about this data to
statistics@dhsspsni.gov.uk

Send questions about privacy to
statistics@dhsspsni.gov.uk

Find out how to suggest improvements to publication at
http://www.dhsspsni.gov.uk/index/stats_research/hospital-stats/...

Open Data Institute, 65 Clifton Street, London EC2A 4JE
info@theodi.org · Company 08030289 · VAT 143 7796 80

Query and Update RDF: (SPARQL)

Photo credit "reedster", Flickr

RDF data access

- The LDP approach is great to build up simple applications
- But as data grows, complexity of relationships grows
  - more sophisticated query possibilities are necessary
  - triples are stored in (possibly very large) databases (“triple stores”)
- How do I query the RDF data?

Querying RDF graphs

- Remember the Python+RDFLib idiom:

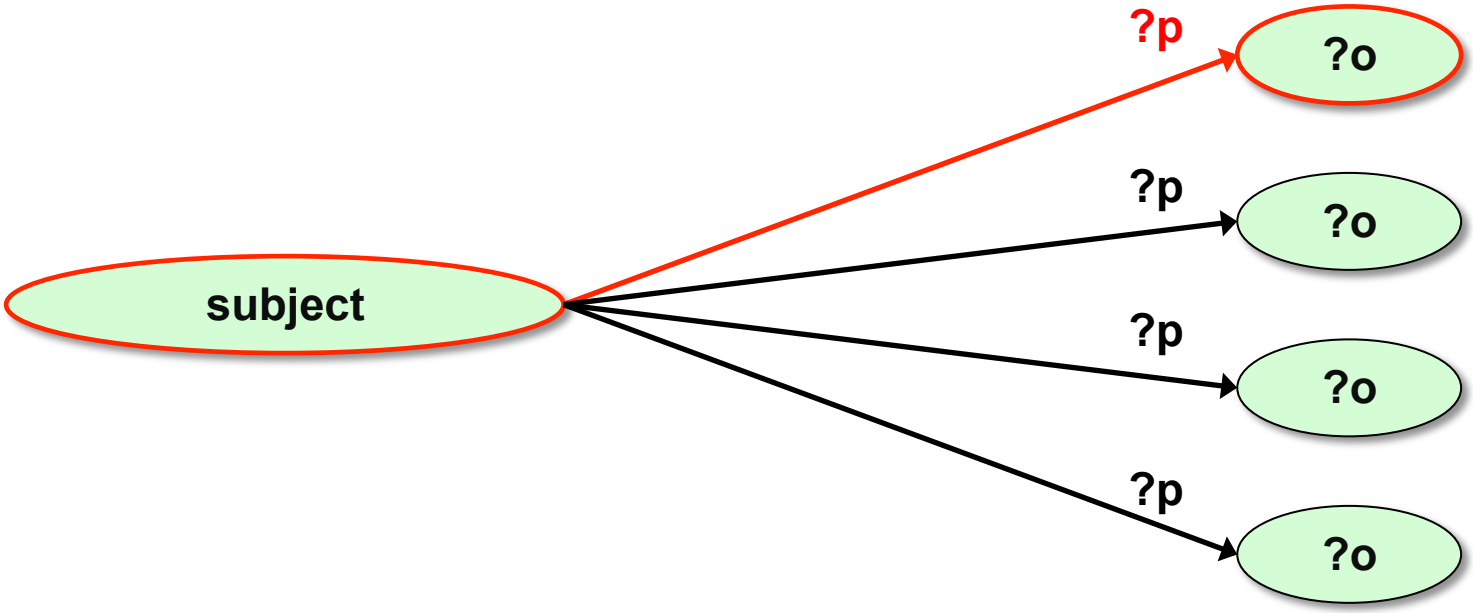
```
for (s,p,o) in graph.triples((subject,None,None)) :
  do_something(p,o)
```

Querying RDF graphs

- In practice, more complex queries into the RDF data are necessary
  - something like: “give me the (a, b) pair of resources, for which there is an x such that (x parent a) and (b brother x) holds” (i.e., return the uncles)
    - these rules may become quite complex
- The goal of SPARQL (Query Language for RDF)

Analyze the Python+RDFLib example

```
for (s,p,o) in graph.triples((subject,None,None)) :
  do_something(p,o)
```



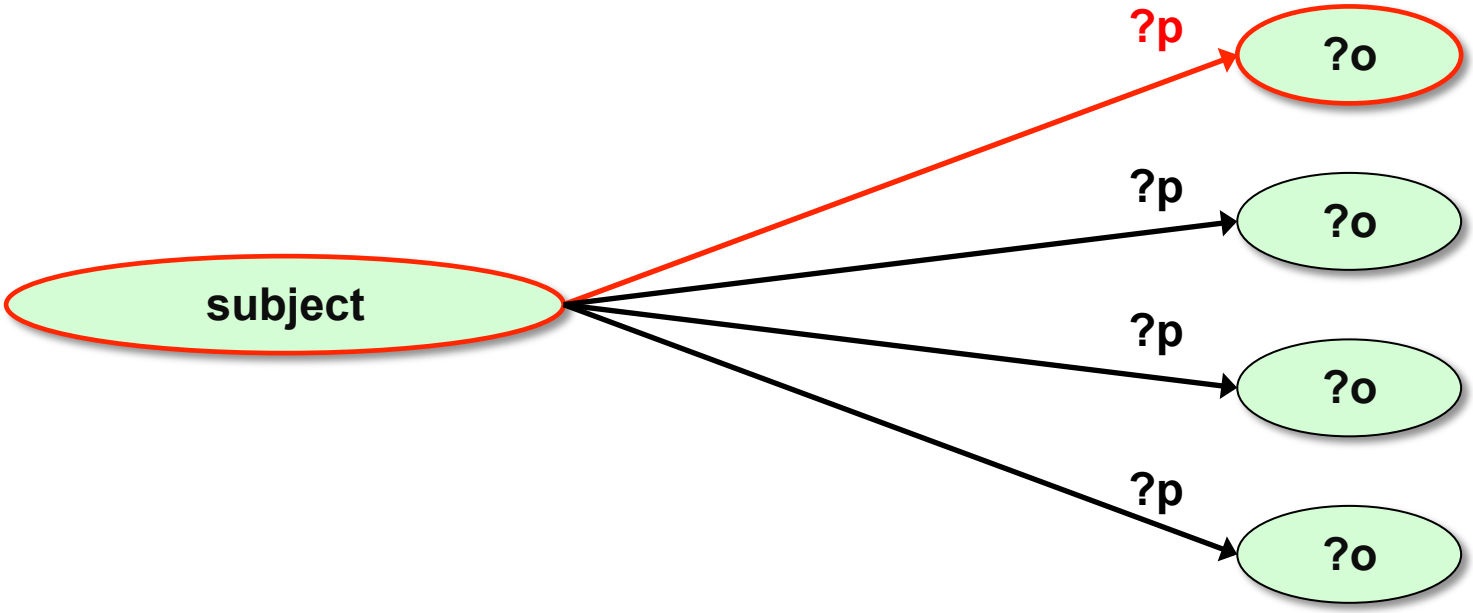
General: graph patterns

- The fundamental idea: use graph patterns
  - the pattern contains unbound symbols
  - by binding the symbols, subgraphs of the RDF graph are selected
  - if there is such a selection, the query returns the bound resources

Our Python example in SPARQL

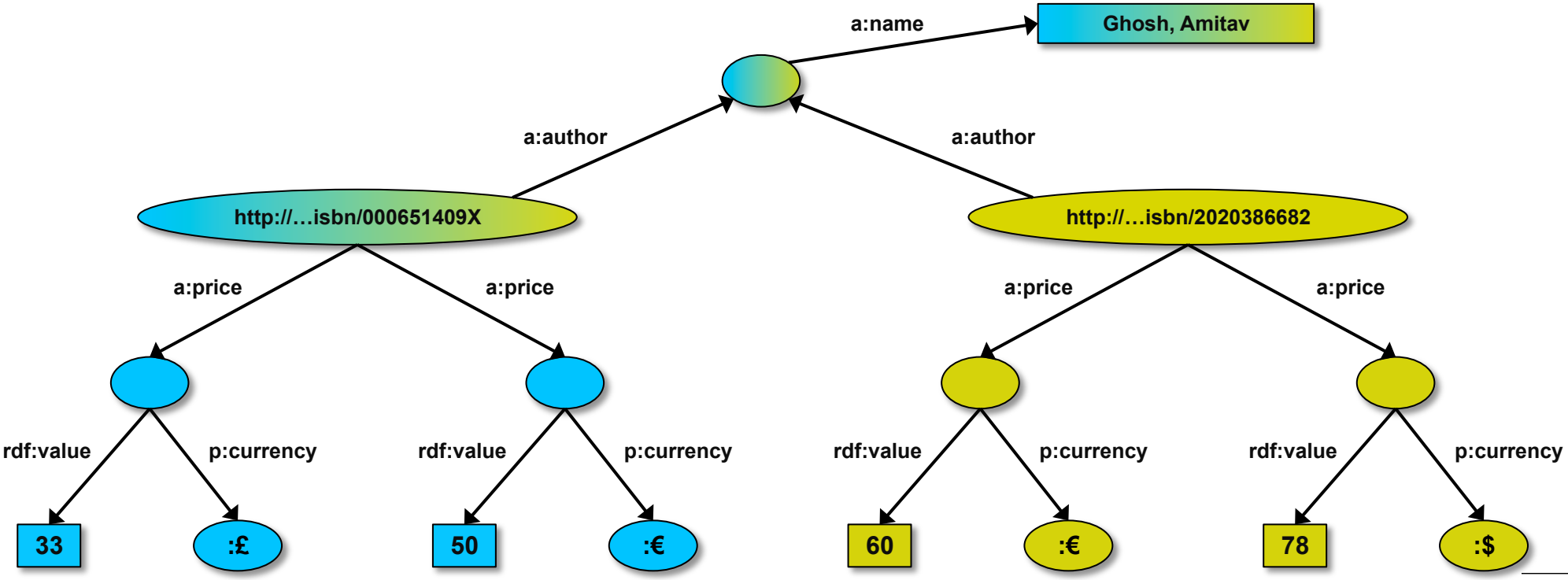
```
SELECT ?p ?o
WHERE {subject ?p ?o}
```

- The triples in WHERE define the graph pattern, with ?p and ?o “unbound” symbols
- The query returns all p, o pairs



Simple SPARQL example

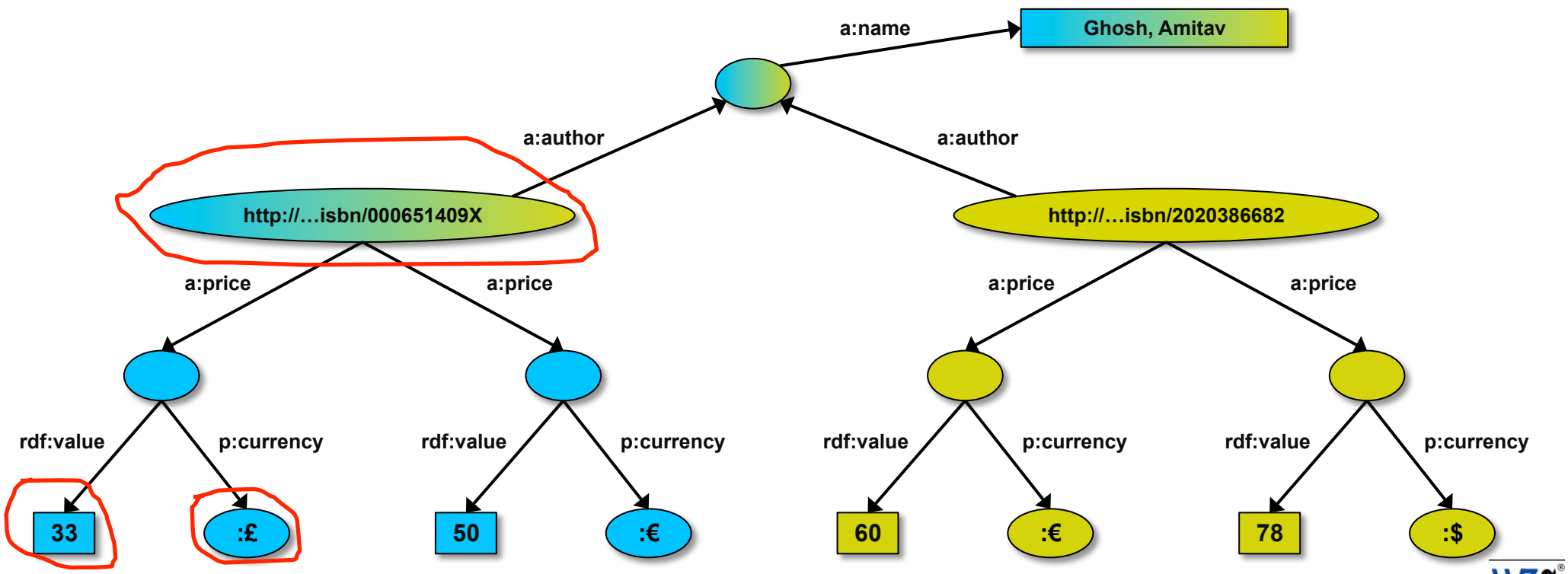
```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```



Simple SPARQL example

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```

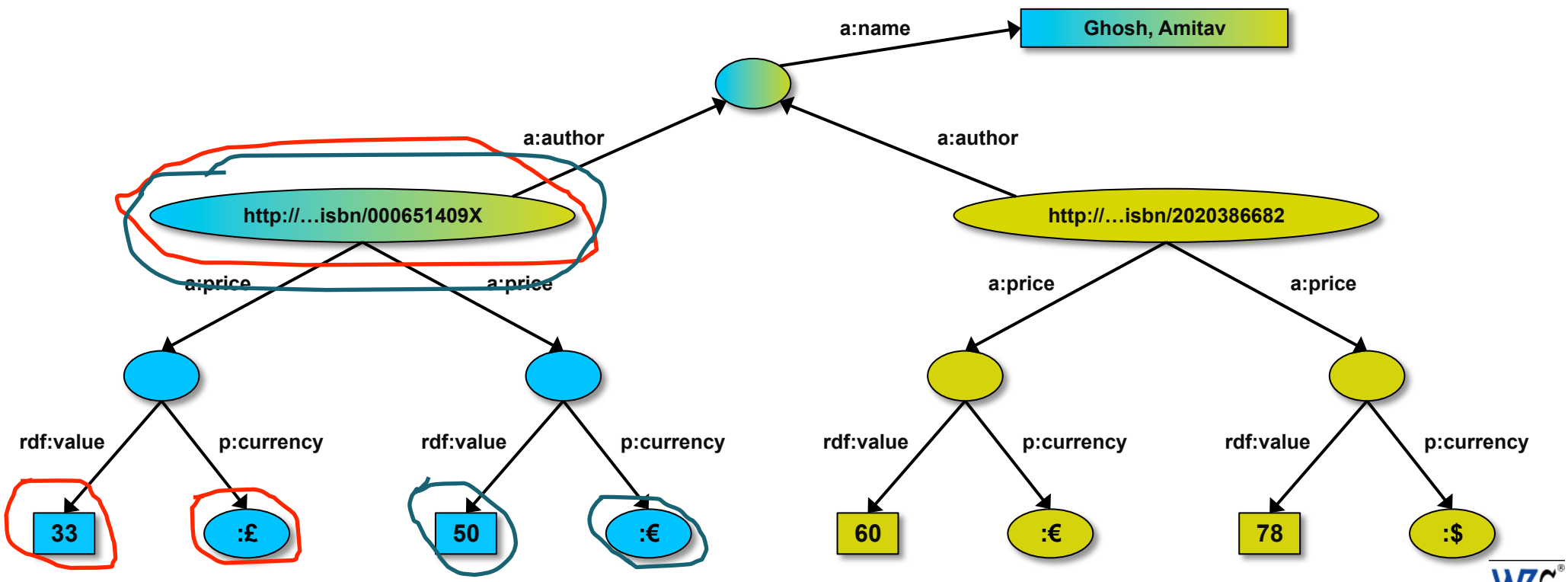
Returns: [<…409X>,33,:£]



Simple SPARQL example

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```

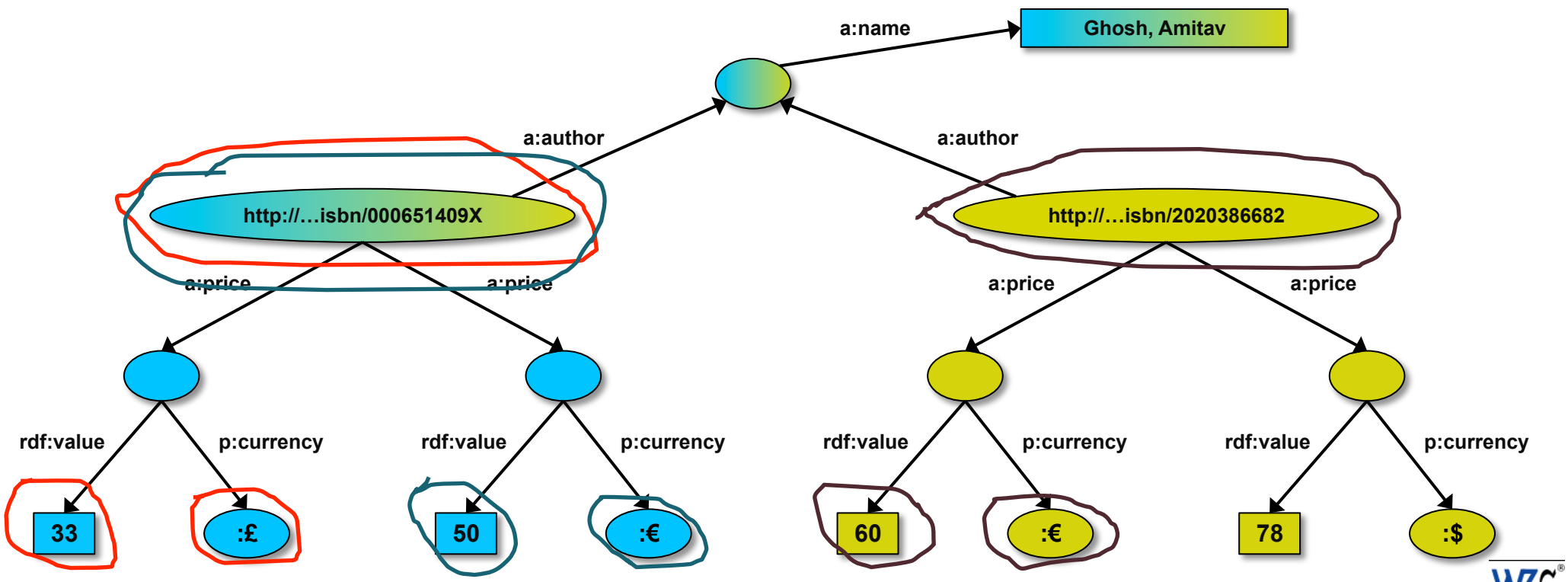
Returns: [<…409X>,33,:£], [<…409X>,50,:€]



Simple SPARQL example

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```

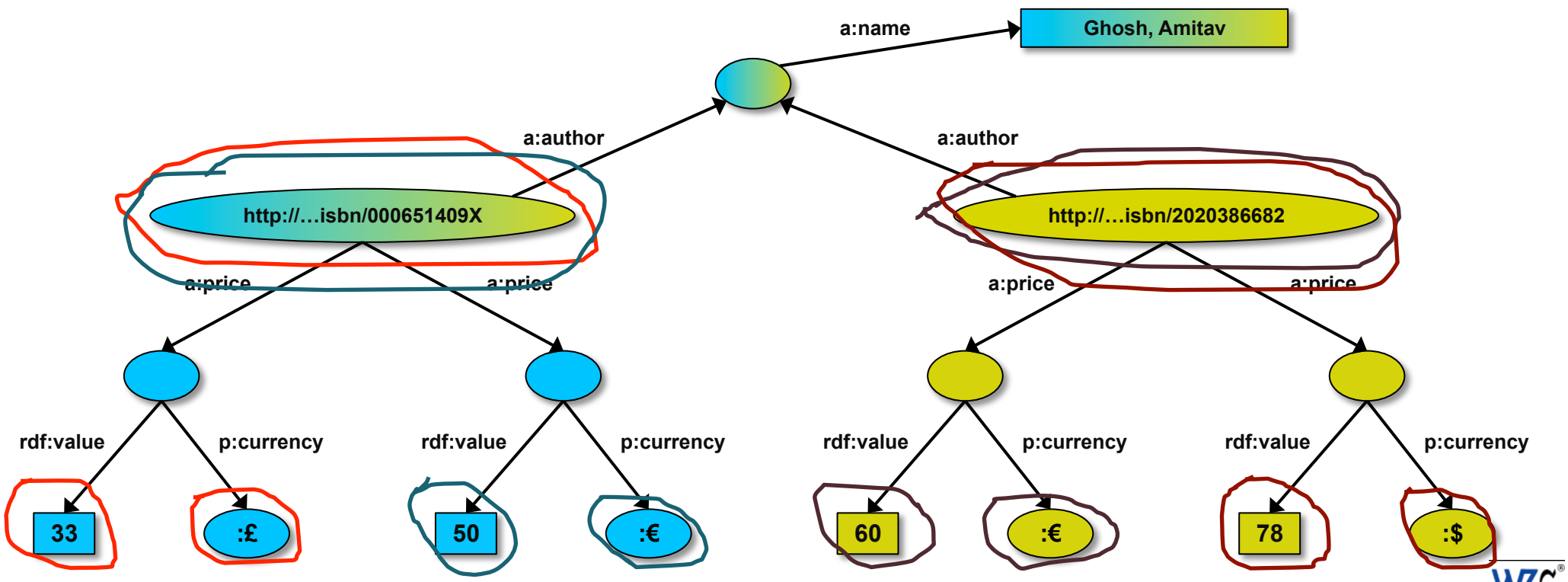
Returns: [<…409X>,33,:£], [<…409X>,50,:€],
[<…6682>,60,:€]



Simple SPARQL example

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE {?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.}
```

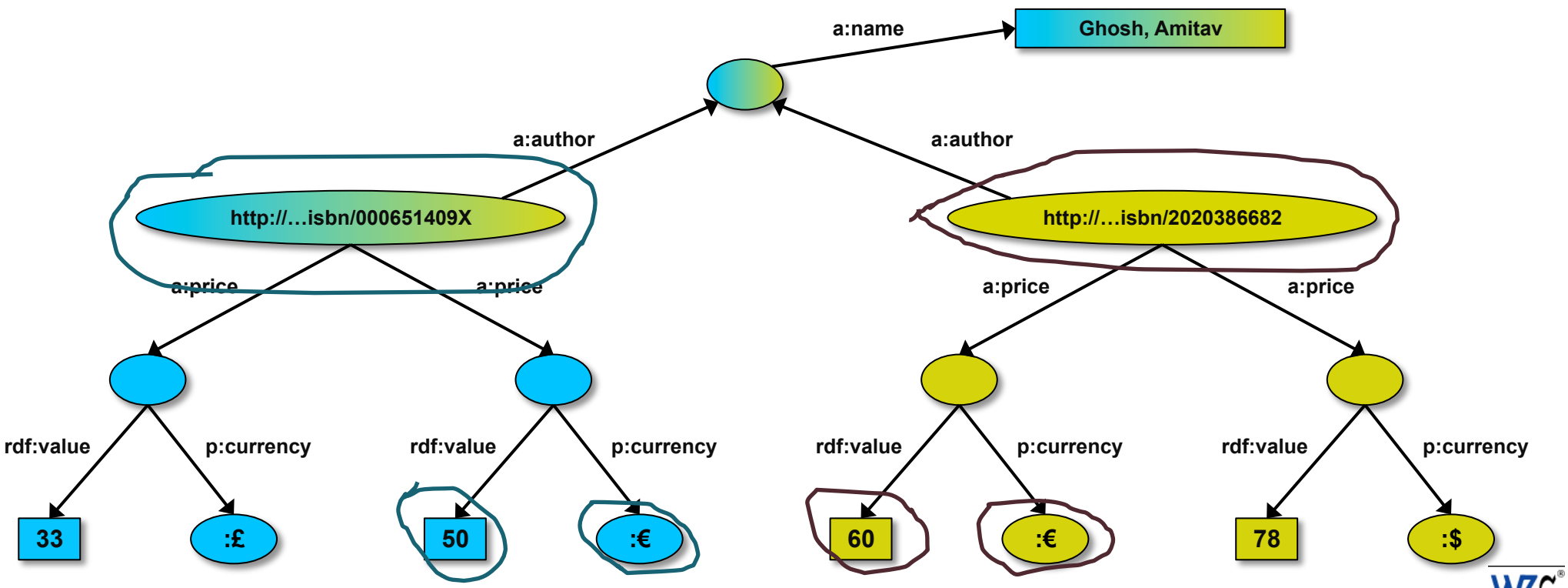
Returns: [<…409X>,33,:£], [<…409X>,50,:€],
[<…6682>,60,:€], [<…6682>,78,:$]



Pattern constraints

```
SELECT ?isbn ?price ?currency # note: not ?x!
WHERE { ?isbn a:price ?x. ?x rdf:value ?price. ?x p:currency ?currency.
        FILTER(?currency == :€) }
```

Returns: [<…409X>,50,:€], [<…6682>,60,:€]



Other SPARQL features

- Limit the number of returned results; remove duplicates, sort them, …
- Optional patterns
- CONSTRUCT new graphs, not only return data
- Use datatypes and/or language tags when matching a pattern
- Aggregation of the results (min, max, average, etc.)
- Path expressions (a bit like regular expressions)

SPARQL usage in practice

- SPARQL is usually used over the network
  - HTTP request is sent to a SPARQL endpoint
  - return is the result of the SELECT, the CONSTRUCT,…
- Separate documents define the protocol and the result format
    - SPARQL Protocol for RDF with HTTP and SOAP bindings
    - SPARQL results in XML or JSON formats
- Many big datasets usually offer “SPARQL endpoints” using this protocol

SPARQL Update

- SPARQL CONSTRUCT returns a new, modified graph
  - the original data remains unchanged!
- SPARQL Update *modifies the original dataset!*

Bathing Water Data Explorer

English

Bathing Water Data Explorer

Environment Agency

home help

Enter bathing water name, county, district or postcode

Search

Throughout the summer, The Environment Agency of England and Wales collects detailed scientific data on the **cleanliness of our bathing waters** (mostly beaches). Weekly and annual compliance ratings are given:

- meets the **higher** standard
- meets the **minimum** standard
- **fails** to meet the minimum standard

Use this web site to **search for** a bathing water by name, county or postcode, or by **browsing around the map**. You can then see the detailed data the agency has collected about that site.

Key: most recent **weekly classification**:

higher standard, minimum standard, fail, closed

monitoring suspended due to abnormal situation

Bathing Water Data Explorer

English

Bathing Water Data Explorer

Environment Agency

Bathing water at Felixstowe South

home | bathing water information | sampling data | suspensions | help

Suffolk, England **Local authority**: Suffolk Coastal District
Year of designation: 1988

Tweet +1

Felixstowe is a resort town but is also home to the country's largest container port on the Orwell Estuary. Felixstowe South is a sand and shingle beach gently sloping to the sea in a sheltered bay. There is sand at low tide. The beach is backed by a promenade and gardens.

Water sampling point location
lat/long: 51.9543304067933, 1.34146436108153; easting/northing: 629700, 233700
View in: Google Maps, Bing Maps, OpenStreetMap

Nearby bathing waters:
Felixstowe North, Dovercourt, Walton, Frinton

Bathing water quality results

Recent results from water quality assessments under the Bathing Water Directive appear below, or view more details of assessment results

Annual Compliance Results					Latest weekly in-season
2008	2009	2010	2011	2012	Tue, 25 Jun 2013 10:15:00
Higher	Higher	Higher	Higher	Higher	

Catchment map

Powered by Ordnance Survey

© Crown copyright and database rights 2013 Ordnance Survey

EA bathing water sampling location
Surface water catchment boundary

Bathing water map

Powered by Ordnance Survey

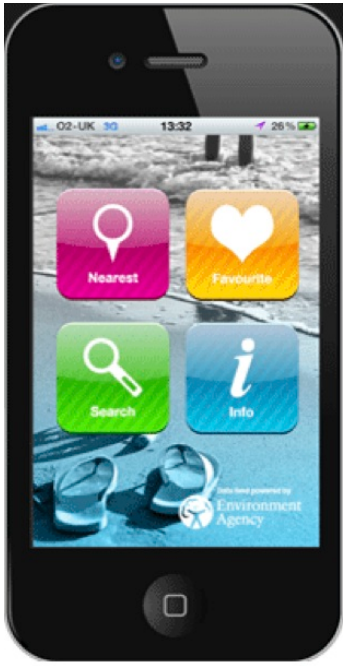
© Crown copyright and database rights 2013 Ordnance Survey

EA bathing water sampling location:
higher standard
minimum standard
fail
closed
not classified

Surface water outfall
Emergency or storm overflow
River or stream
Treated sewage works outfall
Disused outfall

Catchment description

The River Deben is 5km to the north and this drains a large mixed catchment. The Orwell and Stour estuary is 3km to the south and drains a large catchment containing some industry and the Port of Felixstowe



Vocabularies

Photo credit "Indy Reading Coalition", Wordpress.com

But what about RDFS?

- RDFS is enough for many vocabularies
- But not for all!

Some technologies have emerged

- To re-use thesauri, glossaries, etc: SKOS
- To define more complex vocabularies with a strong logical underpinning: OWL

Thesauri, glossaries (SKOS)

Photo credit "scarletgreen", Flickr

SKOS

- Represent and share classifications, glossaries, thesauri, etc
  - for example:
    - Dewey Decimal Classification, Art and Architecture Thesaurus, ACM classification of keywords and terms…
    - classification/formalization of Web 2.0 type tags
- Define classes and properties to add those structures to an RDF universe
  - allow for a quick port of this traditional data, combine it with other data

The term “Fiction”, as defined by the Library of Congress

Fiction – LC Linked Data Se

id.loc.gov/authorities/subjects/sh85048050.html

Private webmail | 2013 | 2012 | LocalData | TR | TR List | Private | Social | Mailing lists | SW | Python | RDFa it! | » | Other Bookmarks

From Library of Congress Subject Headings

Details | Visualization | Suggest Terminology

Fiction

URI(s)
- http://id.loc.gov/authorities/subjects/sh85048050
- info:lc/authorities/sh85048050
- http://id.loc.gov/authorities/sh85048050#concept

Instance Of
- MADS/RDF Topic
- MADS/RDF Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Subject Headings

Collection Membership(s)
- LCSH Collection - Authorized Headings
- LCSH Collection - General Collection

Variants
- Fiction--Philosophy
- Metafiction
- Novellas (Short novels)
- Novels
- Stories

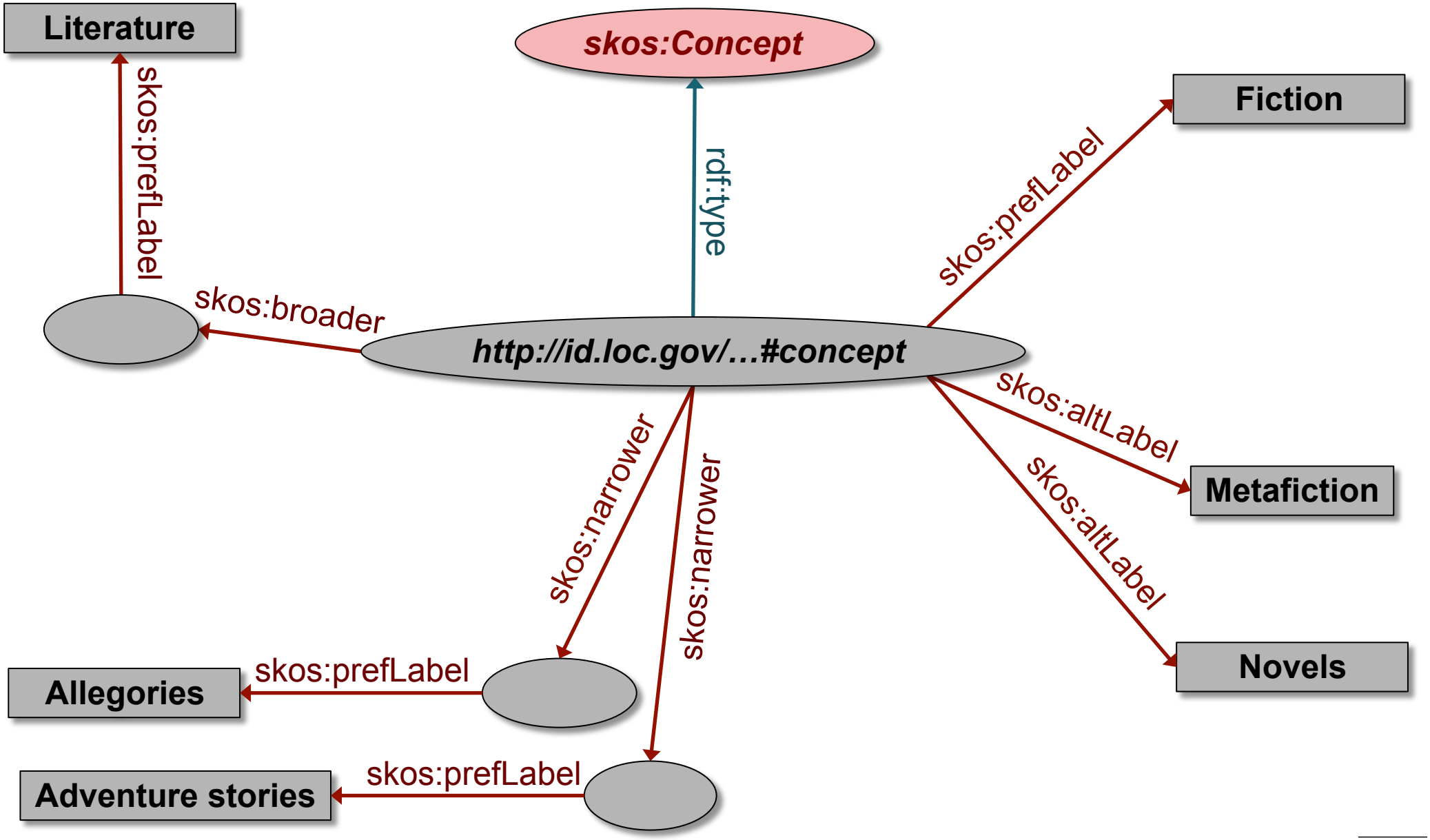
Broader Terms
- Literature

Narrower Terms
- Adventure stories
- Allegories

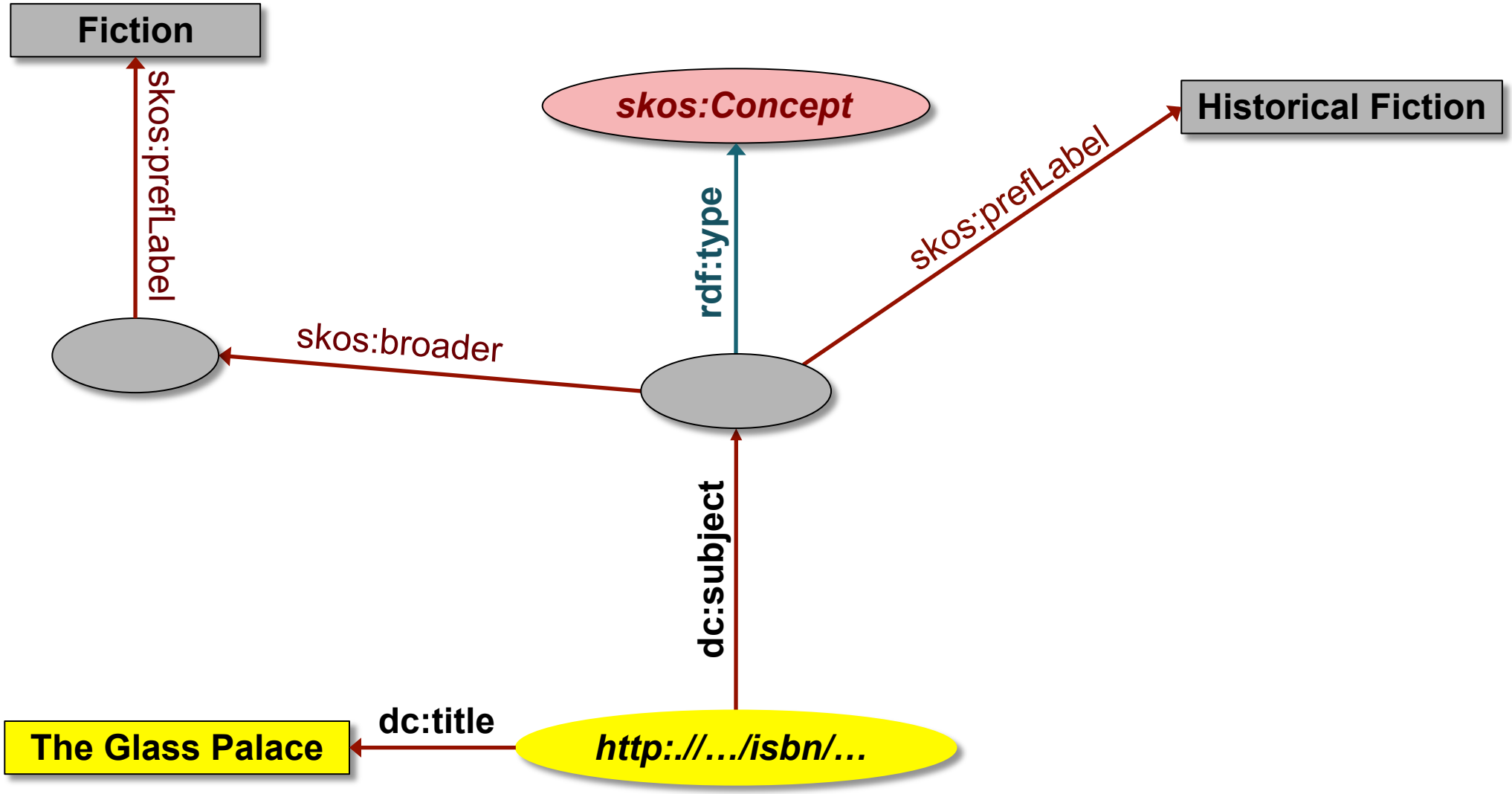
Thesauri have identical structures…

- The structure of the LOC page is fairly typical
  - label, variants/alternate label, narrower, broader, …
  - there is even an ISO standard for these
- SKOS provides a basic structure to create an RDF representation of these

LOC's "Fiction" in SKOS/RDF



Usage of the LOC graph



Importance of SKOS

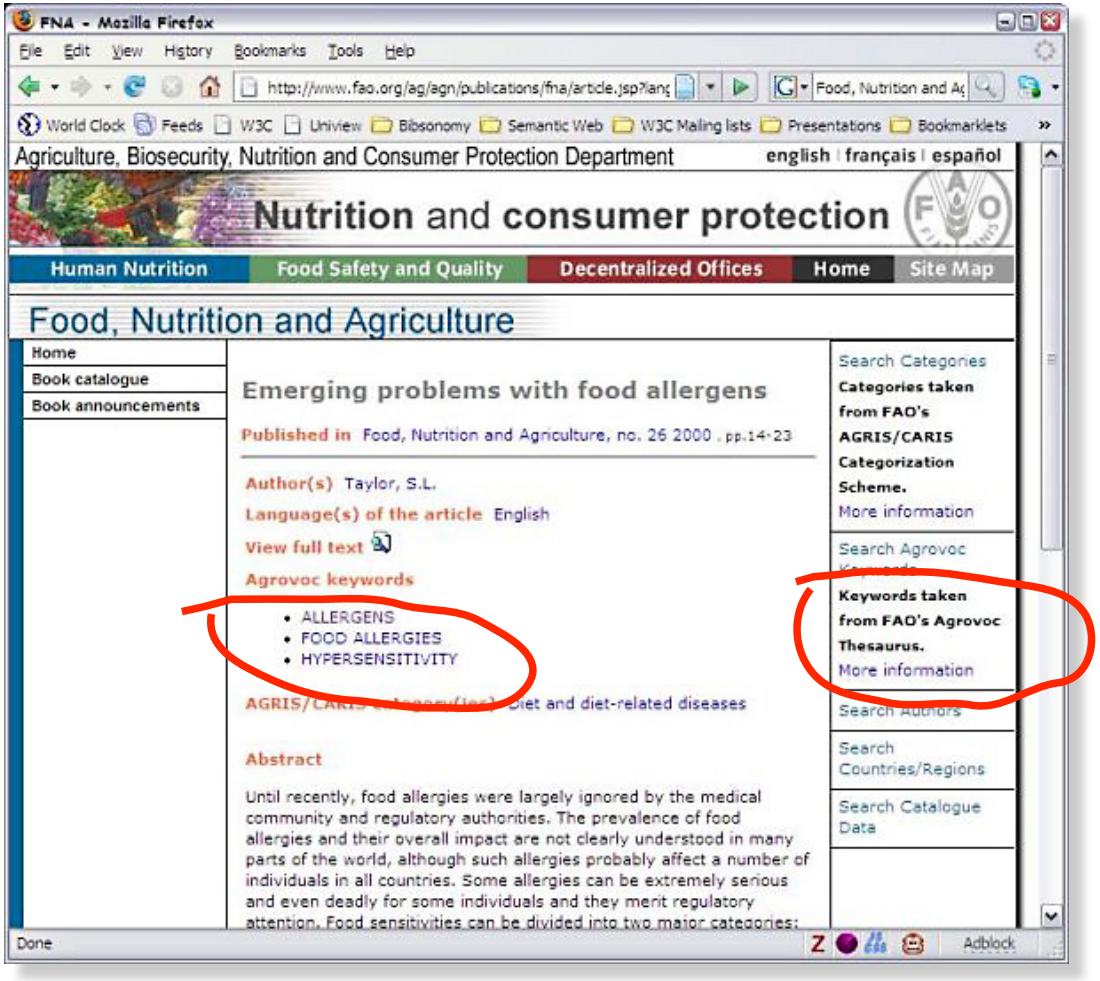
- SKOS provides a simple bridge between the “print world” and the (Semantic) Web
- Thesauri, glossaries, etc., from the library community can be made available
  - LOC is a good example
- SKOS can also be used to organize, e.g., tags, annotate other vocabularies, …

Importance of SKOS

- Anybody in the World can refer to common concepts
  - they mean the same for everybody
- Applications may exploit the relationships among concepts
  - e.g., I can make SPARQL searches for, say, works that are categorized as “Fiction”, albeit indirectly only

Example: FAO Journal portal

- Improved search on journal content based on an agricultural ontology and thesaurus (AGROVOC)



(165) *Courtesy of Gauri Salokhe, Margherita Sini, and Johannes Keizer, FAO, (SWEO Case Study)*

Ontologies (OWL)

Photo credit "Rickydavid", Flickr

SKOS is not enough…

- SKOS may be used to provide simple vocabularies
- But it is not a complete solution
  - it concentrates on the concepts only
  - no characterization of properties in general
  - simple from a logical perspective
    - i.e., only a few inferences are possible

Application may want more…

- Complex applications may want more possibilities:
  - characterization of properties
  - identification of objects with different URI-s
  - disjointness or equivalence of classes
  - construct classes, not only name them
  - more complex classification schemes
  - can a program reason about some terms? E.g.:
    - “if «Person» resources «A» and «B» have the same «foaf:email» property, then «A» and «B» are identical”
  - etc.

Web Ontology Language = OWL

- OWL is an extra layer, a bit like RDF Schemas
  - own namespace, own terms
  - it relies on RDF Schemas
- It is a separate recommendation

OWL is complex…

- OWL is a large set of additional terms
- We will not cover the whole thing here…

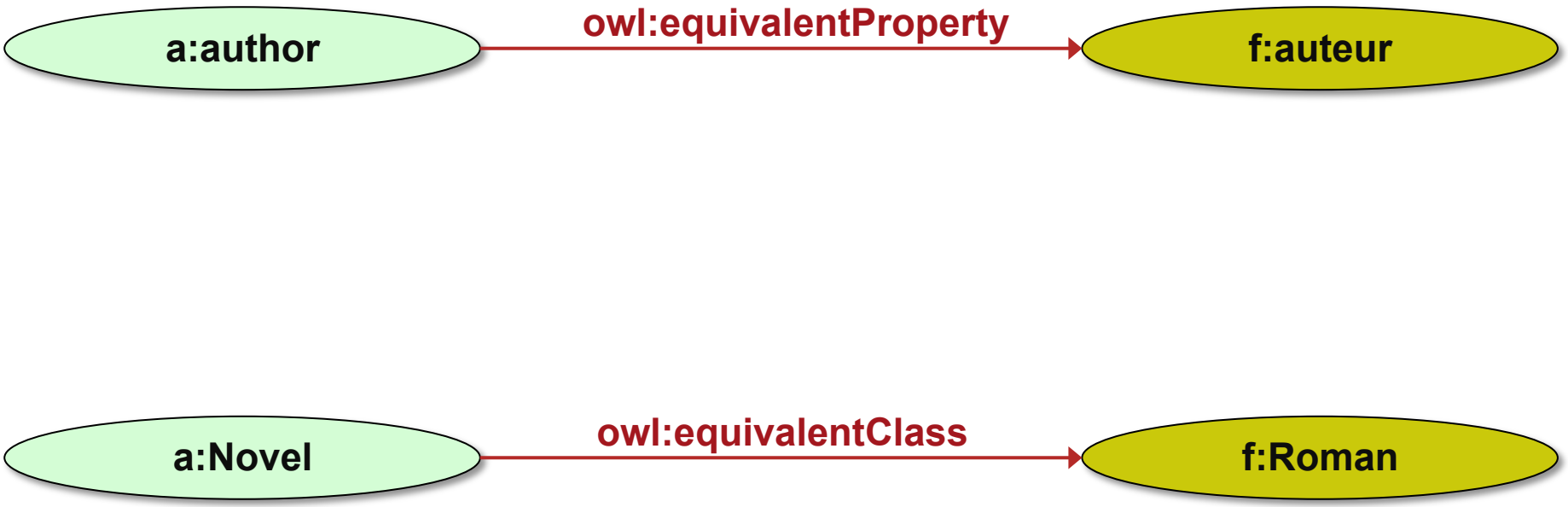
Term equivalences

- For classes:
  - owl:equivalentClass: two classes have the same individuals
  - owl:disjointWith: no individuals in common
- For properties:
  - owl:equivalentProperty
    - remember the a:author vs. f:auteur?
  - owl:propertyDisjointWith

Term equivalences

- For individuals:
  - owl:sameAs: two URIs refer to the same concept (“individual”)
  - owl:differentFrom: negation of owl:sameAs

Connecting to French



Typical usage of owl:sameAs

- Linking our example of Kolkata from one data set (DBpedia) to the other (Geonames):

```
<http://dbpedia.org/resource/Kolkata>
  owl:sameAs <http://sws.geonames.org/1275004>;
```

- This is a major mechanism of “Linking” in the Linked Open Data project

Property characterization

- In OWL, one can characterize the behavior of properties (symmetric, transitive, functional, inverse functional, reflexive, irreflexive, …)
- OWL also separates data and object properties
  - “datatype property” means that its range are typed literals

What this means is…

- If the following holds in our triples:

```
:email rdf:type owl:InverseFunctionalProperty.
```

What this means is…

- If the following holds in our triples:

```
:email rdf:type owl:InverseFunctionalProperty.
<A> :email "mailto:a@b.c".
<B> :email "mailto:a@b.c".
```

What this means is…

- If the following holds in our triples:

```
:email rdf:type owl:InverseFunctionalProperty.
<A> :email "mailto:a@b.c".
<B> :email "mailto:a@b.c".
```

then, processed through OWL, the following holds, too:

```
<A> owl:sameAs <B>.
```

Classes in OWL

- In RDFS, you can subclass existing classes… that's all
- In OWL, you can construct classes from existing ones:
  - enumerate its content
  - through intersection, union, complement
  - etc.

Enumerate class content

```
:Currency
    rdf:type owl:Class;
    owl:oneOf (:€ :£ :$).
```

- I.e., the class consists of *exactly* of those individuals and nothing else

Union of classes

```
:Novel            rdf:type owl:Class.
:Short_Story      rdf:type owl:Class.
:Poetry           rdf:type owl:Class.
:Literature rdf:type owl:Class;
   owl:unionOf (:Novel :Short_Story :Poetry).
```

- Other possibilities: complementOf, intersectionOf, …

For example…

If:

```
:Novel            rdf:type owl:Class.
:Short_Story      rdf:type owl:Class.
:Poetry           rdf:type owl:Class.
:Literature rdf:type owl:Class;
   owl:unionOf (:Novel :Short_Story :Poetry).

<myWork> rdf:type :Novel .
```

then the following holds, too:

```
<myWork> rdf:type :Literature .
```

It can be a bit more complicated…

If:

```
:Novel            rdf:type owl:Class.
:Short_Story      rdf:type owl:Class.
:Poetry           rdf:type owl:Class.
:Literature rdf:type owlClass;
   owl:unionOf (:Novel :Short_Story :Poetry).

fr:Roman owl:equivalentClass :Novel .

<myWork> rdf:type fr:Roman .
```

then, through the combination of different terms, the following still holds:

```
<myWork> rdf:type :Literature .
```

What we have so far…

- The OWL features listed so far are already fairly powerful
  - e.g., various databases can be linked via owl:sameAs, functional or inverse functional properties, etc.
- Many inferred relationship can be found using a traditional rule engine

However… that may not be enough

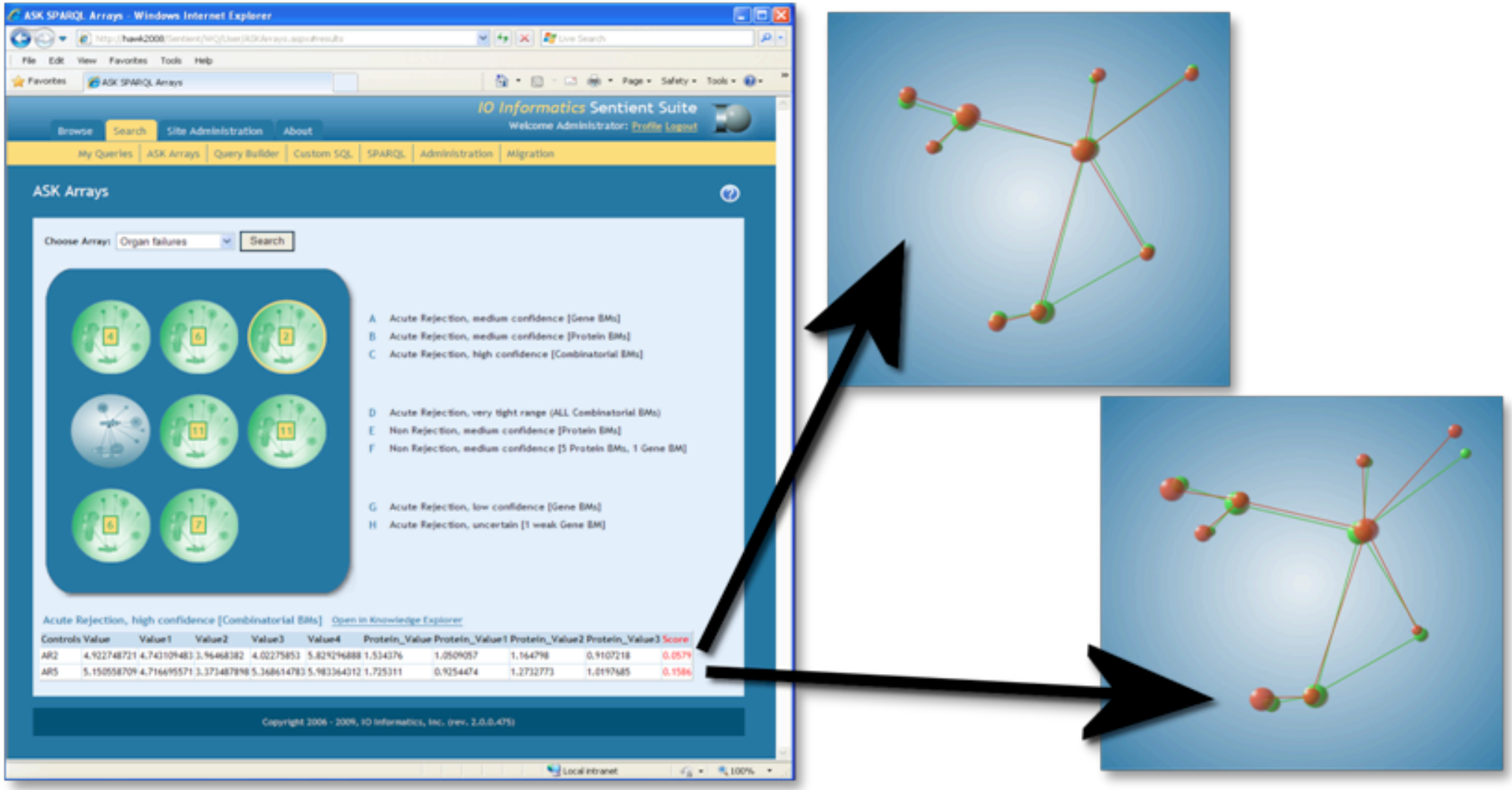
- Very large vocabularies might require even more complex features
  - typical usage example: definition of all concepts in a health care environment
  - some major issues
    - the way classes (i.e., “concepts”) are defined
    - handling of datatypes
- OWL includes those extra features but… the inference engines become (much) more complex ☹

Example: Organ Failure Risk Detection

- System by IO Informatics and UBC:
  - data integrated from experimental data, clinical endpoints, public ontologies, LOD, etc.
  - statistical analysis is performed on the data
  - SPARQL is used to query the results
    - a visual interface is provided
    - for clinicians, a simple web-based alerting of “hits” is provided with statistical scores

(186) *Courtesy of Robert Stanley, et al, IO Informatics, USA, and UBC, Canada, (SWEO Case Study)*

Example: Organ Failure Risk Detection

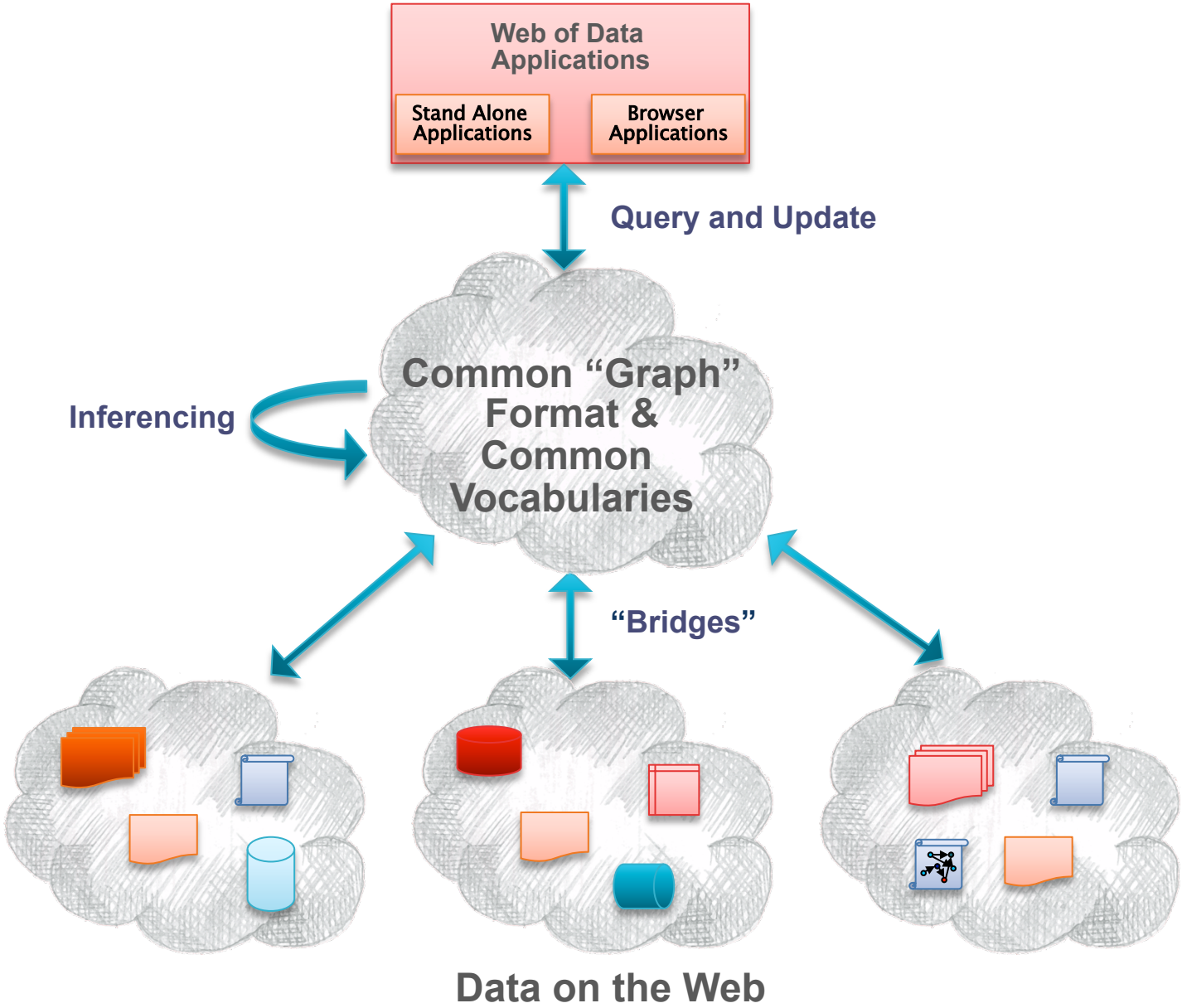


(187) *Courtesy of Robert Stanley, et al, IO Informatics, USA, and UBC, Canada, (SWEO Case Study)*

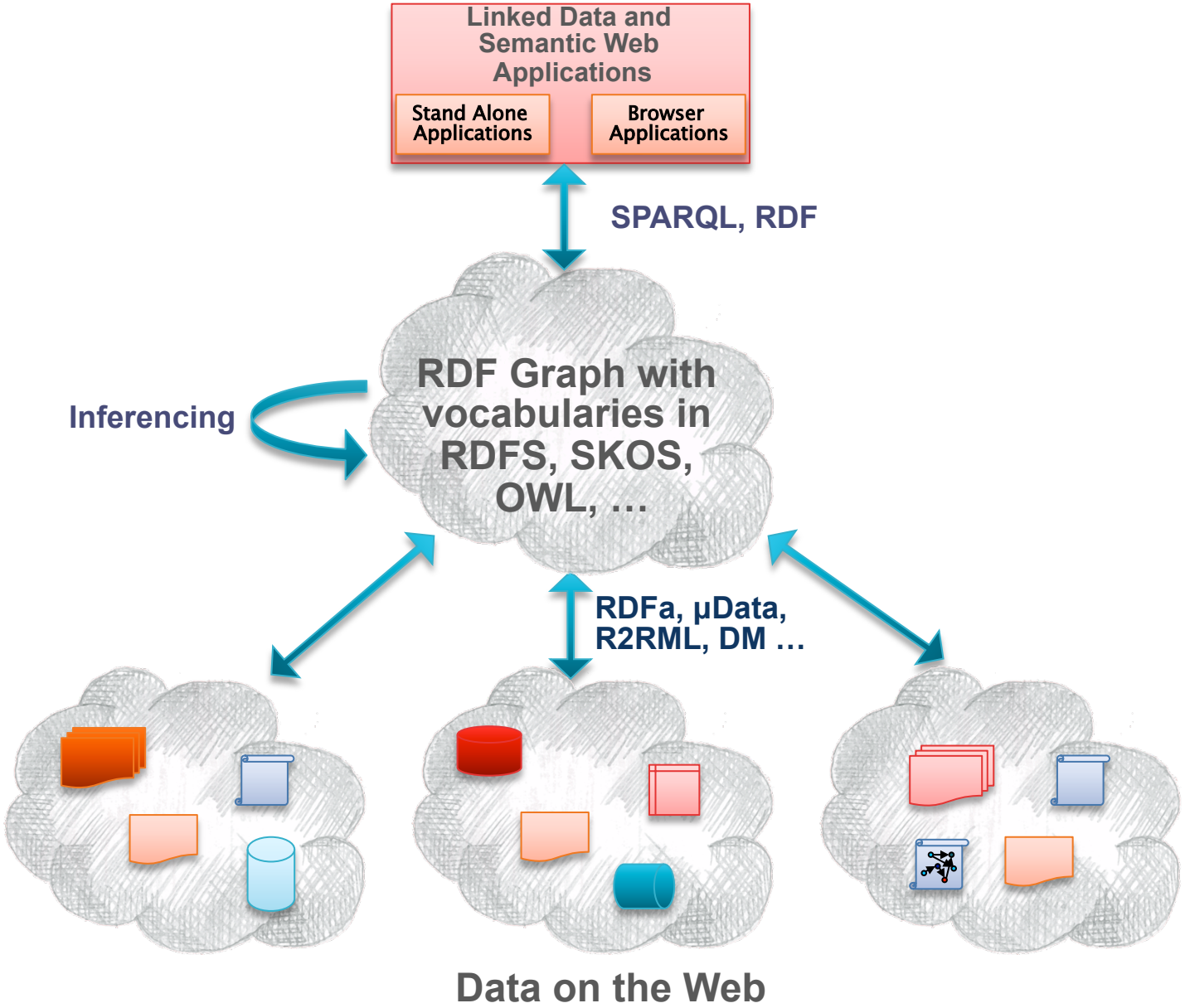
What have we achieved?

(putting all together)

Remember the integration example?



The same with what we learned



Available documents, resources

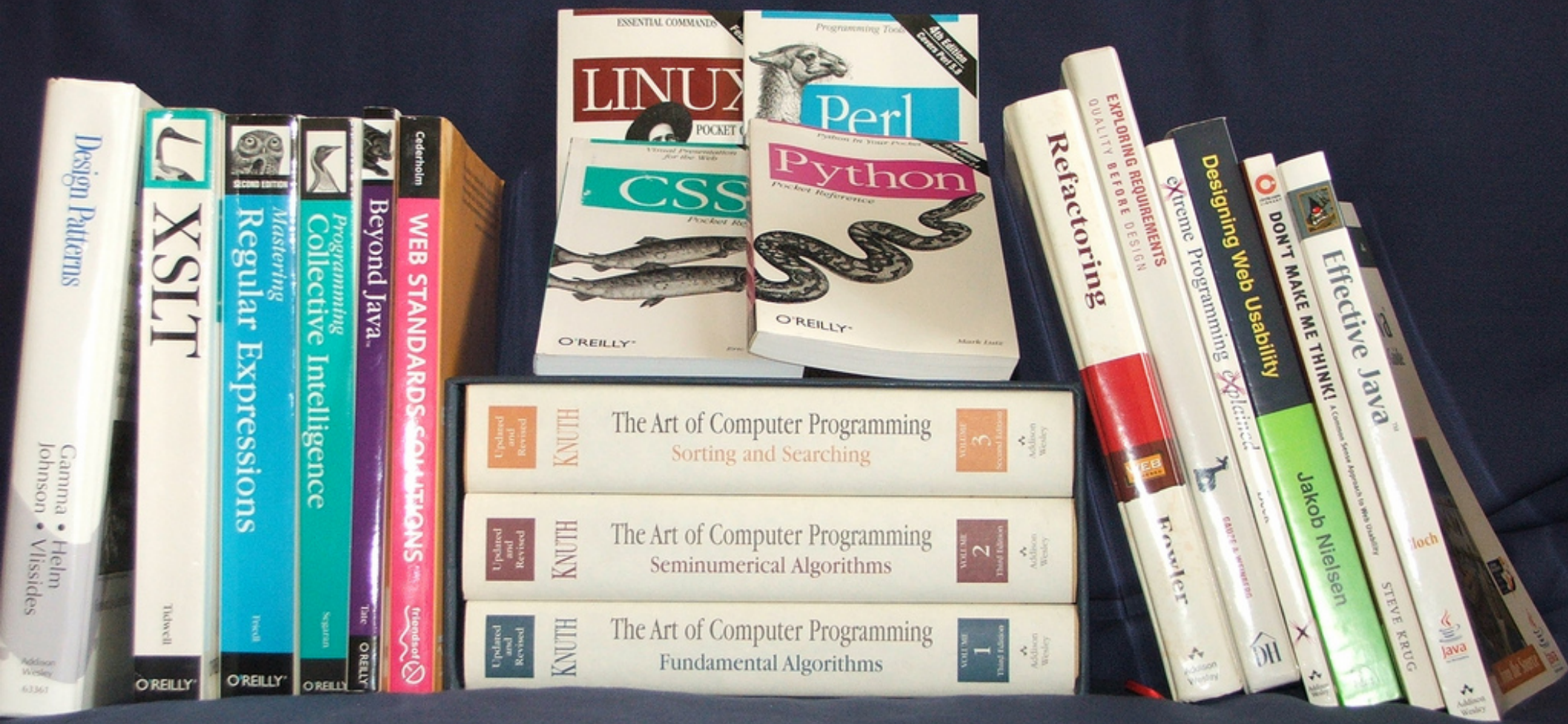


Photo credit "Abizern", Flickr

Available specifications: Primers, Guides

- The “RDF Primer” and the “OWL Guide” give a formal introduction to RDF(S) and OWL
- SKOS has its separate “SKOS Primer”
- GRDDL Primer and RDFa Primer have been published
- The W3C Semantic Web Activity Wiki has links to all the specifications

“Core” vocabularies

- There are also a number “core vocabularies”
  - *Dublin Core*
  - FOAF: about people and their organizations
  - DOAP: on the descriptions of software projects
  - SIOC: Semantically-Interlinked Online Communities
  - vCard in RDF
  - …

Some books

- D. Wood, M. Zaidman, and L. Ruth: *Linked Data*, 2013
- G. Antoniu, P. Groth, F. van Harmelen, and R. Hoekstra: *Semantic Web Primer, 3rd edition,* 2012
- T. Heath and C. Bizer: *Linked Data: Evolving the Web into a Global Data Space*, 2011
- D. Allemang and J. Hendler: *Semantic Web for the Working Ontologist 2nd edition*, 2011
- M. Watson: *Practical Semantic Web and Linked data Applications*, 2010
- P. Hitzler, R. Sebastian, and M. Krötzsch: *Foundation of Semantic Web Technologies*, 2009
- …

See the separate Wiki page collecting book references

Further information

- Planet RDF aggregates a number of SW blogs:
  - http://planetrdf.com/
- Semantic Web Interest Group
  - a forum developers with a publicly archived mailing list, and a constant IRC presence on freenode.net#swig
  - anybody can sign up on the list
    - http://www.w3.org/2001/sw/interest/
- Linked Data mailing list
  - a forum concentrating on linked data with a public archive
  - anybody can sign up on the list
    - http://lists.w3.org/Archives/Public/public-lod/

Lots of Tools (*not* an exhaustive list!)

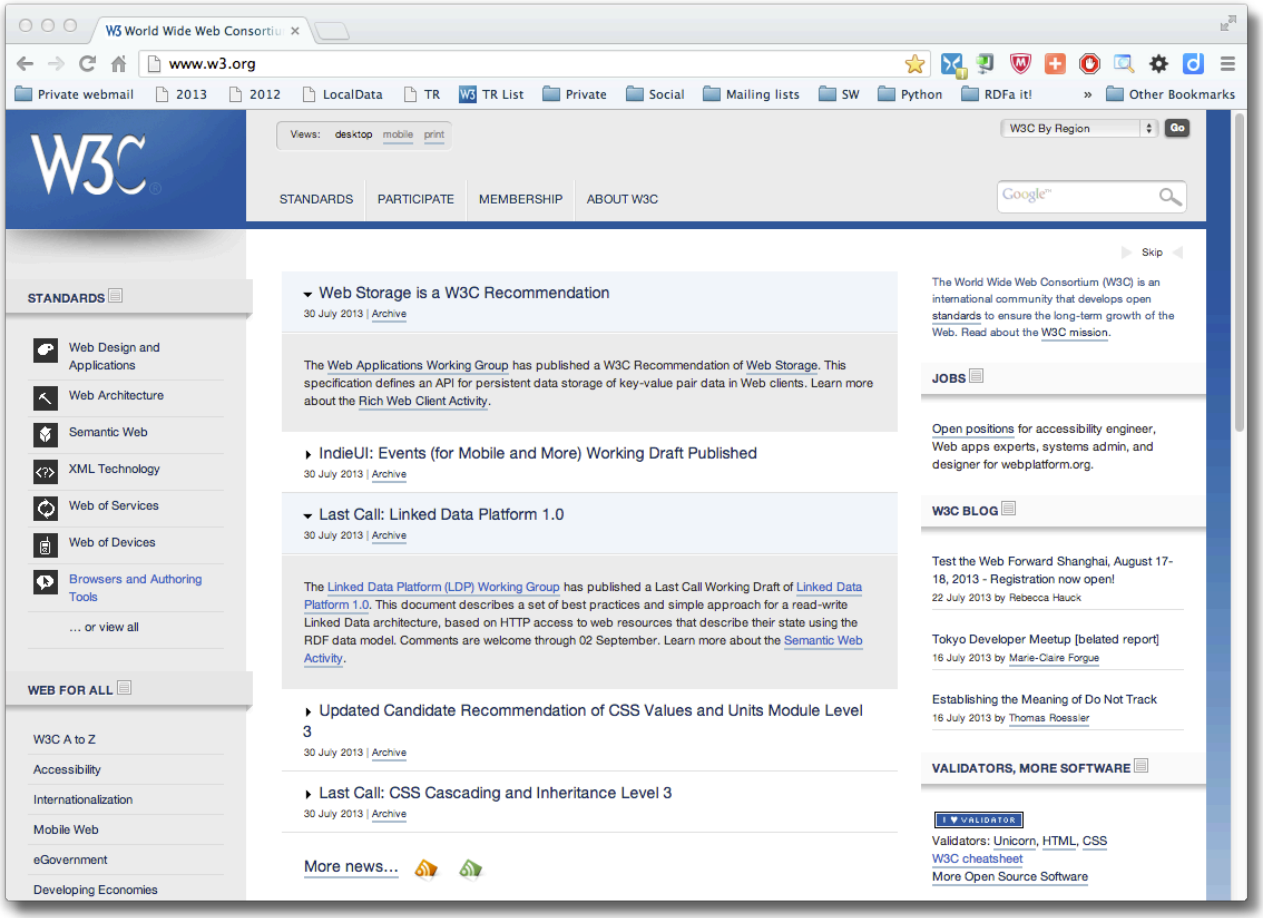
- Some names:
  - Jena, AllegroGraph, Mulgara, Sesame, flickurl, 4Store, …
  - TopBraid Suite, Virtuoso environment, Falcon, Drupal 7, Redland, Pellet, …
  - Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, Talis Platform, …
  - RDF Gateway, RDFLib, Open Anzo, DartGrid, Zitgist, Ontotext, Protégé, …
  - Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore…
  - …
- Categories:
  - Triple Stores
  - Inference engines
  - Converters
  - Search engines
  - Middleware
  - CMS
  - Semantic Web browsers
  - Development environments
  - Semantic Wikis
  - …

More on http://www.w3.org/2001/sw/wiki/Tools

Conclusions

- Linked Data, the Semantic Web, and related technologies are there to integrate data on the Web
- The goal is the creation of a Web of Data

Lot remain to be done…



- Lots of issues to be solved
- But… W3C needs experts!
  - consider joining W3C, as well as the work done there!


ENJOY THE
CONFERENCE!

Thank you for your attention

These slides are also available on the Web:

http://www.w3.org/2013/Talks/0902-Lisbon-IH/